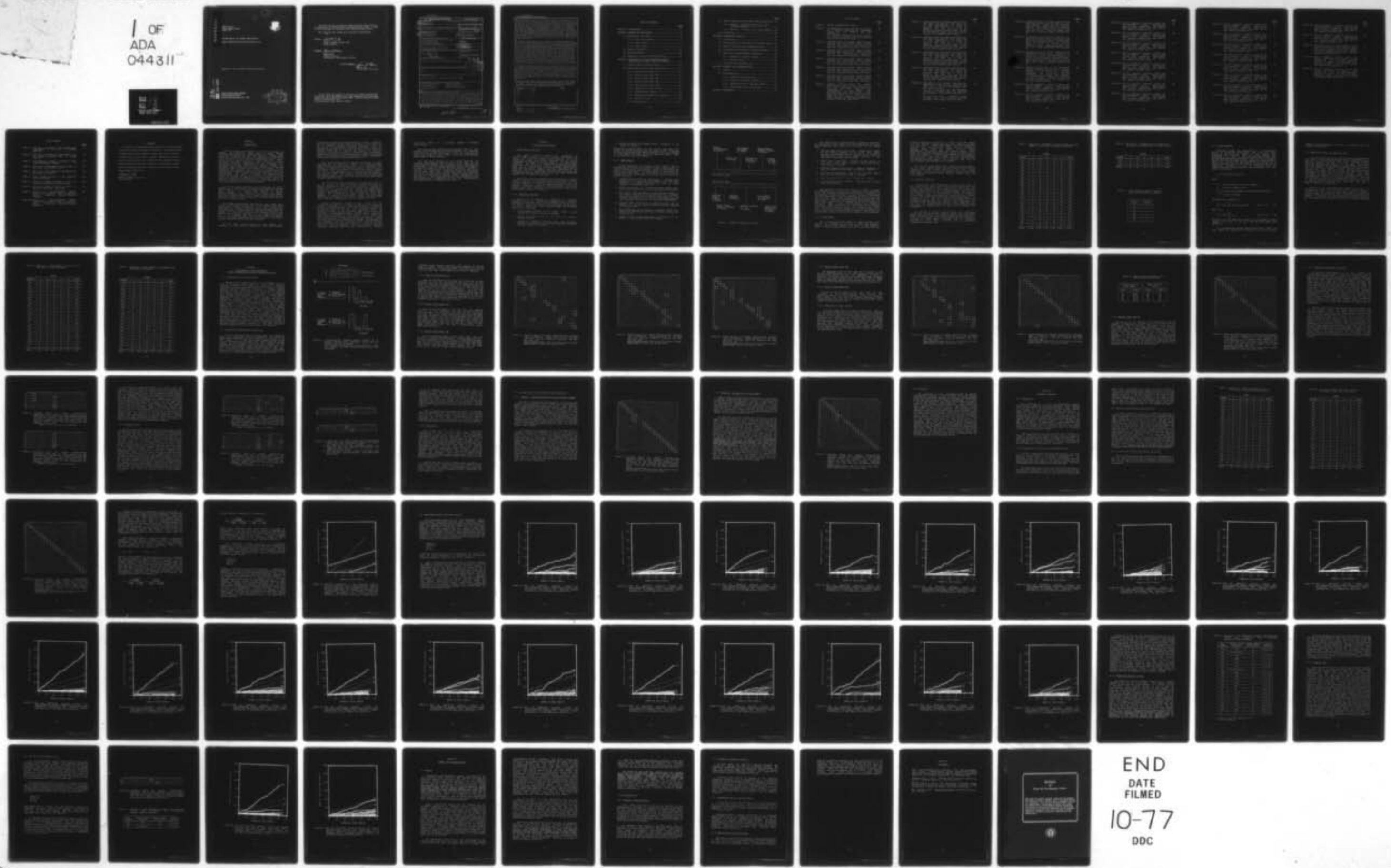
AD-A044 311 SPEECH COMMUNICATIONS RESEARCH LAB INC SANTA BARBARA CALIF F/G 5/7
FEATURE ANALYSIS FOR SPEAKER IDENTIFICATION.(U)
AUG 77 L L PFEIFER F0602-76-C-0157

UNCLASSIFIED RADC-TR-77-277 NL

1 OF
ADA
044311



END
DATE
FILMED
10-77
DDC

12

RADC-TR-77-277
Final Technical Report
August 1977

FEATURE ANALYSIS FOR SPEAKER IDENTIFICATION

Speech Communications Research Laboratory, Inc.

Approved for public release; distribution unlimited.

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public including foreign nations.

This report has been reviewed and is approved for publications.

APPROVED: Robert A Curtis

ROBERT A. CURTIS, Captain, USAF
Project Engineer

APPROVED: Howard Davis

HOWARD DAVIS
Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (DAP) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.
Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

12 **REPORT DOCUMENTATION PAGE**

READ INSTRUCTIONS
BEFORE COMPLETING FORM

~~1. REPORT NUMBER~~ 18 RADC-TR-77-277	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 FEATURE ANALYSIS FOR SPEAKER IDENTIFICATION.		~~5. TYPE OF REPORT & PERIOD COVERED~~ 9 Final Technical Report.
		~~6. PERFORMING ORG. REPORT NUMBER~~ N/A
~~7. AUTHOR(s)~~ 10 Larry L. Pfeifer		8. CONTRACT OR GRANT NUMBER(s) 15 F30602-76-C-0157
9. PERFORMING ORGANIZATION NAME AND ADDRESS Speech Communications Research Laboratory, Inc. 800 A Miramonte Drive Santa Barbara CA 93109		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 31011G 70550725
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAD) Griffiss AFB NY 13441		~~12. REPORT DATE~~ 11 August 1977
		~~13. NUMBER OF PAGES~~ 78
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same 12 93p.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Captain Robert A. Curtis (IRAD)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speaker Identification	Reflection Coefficients
Text-Independent	Sequential Analysis
Inverse Filter Analysis	Weighted Euclidean Distance
Vowel Detection	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A method for text-independent speaker identification has been developed which utilizes vowel sounds as the basis for extracting speaker characteristics. The use of this approach typically requires that vowel samples first be classified according to vowel category, so that vowels of the same category can be compared in the speaker identification process. It has been demonstrated, however that it is only necessary to detect vowel-like sounds in the speech material and that speaker identification performance actually improves when there is no vowel recognition. → next page (Con't)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

387 936

cont.

Another significant outcome of this research was the successful application of sequential analysis to the decision process. Sequential analysis relies on the accumulation of speaker classification results from several vowel samples before making a decision. If a sufficient number of test samples are classified as any one speaker, then a decision can be made, with a certain level of confidence, regarding the identity of the unknown speaker. The method allows acceptance and rejection thresholds to be established with specified error probabilities. This makes the sequential analysis procedure a dynamic process which accumulates and tests vowel samples until a confident decision can be made. Thus, the decision time and the amount of speech material needed is variable, depending on the speaker being tested. The sequential analysis is similar to the human perception process where we can quickly identify a unique voice, but listen longer when there is uncertainty.

The experiments which led to the conclusions of the research were based upon a data base of vowel samples from more than one hour of speech material excerpted from conversational speech recordings. Five vowel classes were represented in the data, these being among the most frequently occurring vowel sounds in spoken English. Recordings were made of twenty speakers (ten male and ten female) during conversational interviews. Two of the speakers were recorded two weeks later so that effects of time could be studied as well. A total of 4786 vowels were labelled and analyzed. These data were split into two independent sets for reference samples and test samples. The autocorrelation method of linear prediction was used to analyze a 20 msec window in the steady-state portion of each vowel. Twelve reflection coefficients (k-parameters) were generated and used as feature vectors throughout the study. The first series of three experiments led to two significant results: 1) a majority-rule decision procedure could provide 95 percent correct identification, and 2) it is possible to have high accuracy speaker identification from vowel sounds without vowel recognition.

Additional experimentation involved the use of sequential analysis in the decision process. Decision thresholds were established with the error specifications that P(FR)=0.05 and (FA)=0.05. Using sequential analysis, 18 out of 20 speakers (90 percent) were correctly identified, 17 with the specified probability level, and one at a slightly lower confidence level. The amount of time required to reach a decision ranged between 5.5 seconds and 90 seconds. For one of the test speakers the decision narrowed down to two possible references, with the true speaker being a member of the pair. Only one test speaker in twenty was falsely identified.

Using the same 20-speaker reference patterns, tests with data from two speakers recorded two weeks after the reference material showed that one of the speakers was correctly identified in 18 seconds and the second speaker was correctly identified, but with a lower confidence level than the first.

TABLE OF CONTENTS

PAGE

SECTION 1--INTRODUCTION....................................1

SECTION 2--METHODS AND PROCEDURES..........................4

2.1 Experimental Data Base.................................4

2.1.1 Labelling procedure..................................4

2.1.2 Label format...5

2.1.3 Vowel counts...7

2.2 Analysis Parameters....................................8

2.3 Distance Measure......................................11

2.4 Separation of Test and Reference Data.................12

SECTION 3--DEVELOPMENT OF TEXT-INDEPENDENT SPEAKER IDENTIFICATION WITHOUT VOWEL RECOGNITION..........15

3.1 Confusion Matrix Interpretation.......................15

3.2 Experiments Assuming Vowel Recognition................15

3.2.1 Results using vowel /IY/............................17

3.2.2 Results using vowel /IH/............................17

3.2.3 Results using vowel /EH/............................17

3.2.4 Results using vowel /AE/............................21

3.2.5 Results using vowel /AX/............................21

3.2.6 Comparison of vowel results.........................21

3.2.7 Combined vowel results..............................24

3.2.8 Statistical stability of results....................26

3.2.9 Effects of time.....................................28

3.2.10 Discussion...31

PAGE

3.3 Speaker Identification without Vowel Recognition...32

3.3.1 Method I - Multiple patterns for each reference speaker...32

3.3.2 Method II - Merging of all vowel samples....34

3.4 Discussion...36

SECTION 4--SEQUENTIAL ANALYSIS...37

4.1 Introduction...37

4.2 Reorganization of Data Base into Discourse Order...37

4.3 Method II Results Using Sequential Data...38

4.4 Illustration of Sequential Analysis Procedure......38

4.5 Experiments Using Sequential Analysis...45

4.5.1 Sequential analysis results...66

4.5.2 Special cases...68

4.6 Test Data from Two-Weeks Later...69

SECTION 5--SUMMARY AND RECOMMENDATIONS...73

5.1 Summary...73

5.2 Recommendations...75

5.2.1 Automatic vowel detection...75

5.2.2 Automatic sequential analysis...76

5.2.3 Degradation due to time and channel...76

5.2.4 Identification by verification...76

SECTION 6--REFERENCES...78

LIST OF FIGURES

PAGE

Figure 1. Format of speech-event label. 6

Figure 2. (a) Hypothetical confusion matrix results of a speaker identification experiment having two test speakers and five references.
(b) Histogram interpretation of the results for each test speaker, with each reference being a bin in the histogram. 16

Figure 3. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /IY/ only. Data set 3 used for reference and data set 4 for test. 18

Figure 4. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /IH/ only. Data set 3 used for reference and data set 4 for test. 19

Figure 5. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /EH/ only. Data set 3 used for reference and data set 4 for test. 20

Figure 6. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /AE/ only. Data set 3 used for reference and data set 4 for test. 22

Figure 7. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /AX/ only. Data set 3 used for reference and data set 4 for test. 23

Figure 8. Composite confusion matrix for speaker identification test assuming recognition of five vowel categories. There were five references for each speaker, one corresponding to each vowel class. Reference data taken from data set 3, test data from data set 4. Each input test sample was first classified according to vowel category and then matched with those references for that vowel only. 25

PAGE

Figure 9. Confusion matrix for speaker identification experiment using JDM test data (data set 2), and approximately 1.5 minutes of reference data (data set 3) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results. 27

Figure 10. Confusion matrix for speaker identification experiment using JDM test data (data set 2), and approximately 3.0 minutes of reference data (data set 1) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results. 27

Figure 11. Confusion matrix for speaker identification experiment using BFH test data (data set 2), and approximately 1.5 minutes of reference data (data set 3) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results. 29

Figure 12. Confusion matrix for speaker identification experiment using BFH test data (data set 2), and approximately 3.0 minutes of reference data (data set 1) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results. 29

Figure 13. Comparison of two speaker identification experiments having the same reference set (data set 3) and two test speakers (one male and one female). 30
(a) Test data from the same recording session as reference data, but different time segments. Test samples from data set 4.
(b) Test data from a recording session two-weeks after the reference data was recorded. Test samples from data set 2.

PAGE

Figure 14. Confusion matrix for speaker identification experiment without vowel recognition on the test samples. There were 20 reference speakers, each represented by five patterns (one for each vowel), for a total of 100 reference patterns. Reference samples from data set 3 and test samples from data set 4. 33

Figure 15. Confusion matrix for speaker identification experiment without vowel recognition, but where each reference speaker is represented by the pooled samples of all vowel classes. There is only one pattern for each reference speaker. Reference samples from data set 3 and test samples from data set 4. 35

Figure 16. Confusion matrix for speaker identification experiment without vowel recognition on either the test or reference samples, and the test data presented in discourse order. There were 20 reference patterns, one for each known speaker. Reference samples from data set 5, and test samples from data set 6. 41

Figure 17. Graphical illustration of the sequential analysis decision procedure for an unknown speaker. The two parallel lines represent acceptance and rejection decision thresholds. The stepped lines represent the classification counts for different reference speakers. References A and C are rejected and reference B is accepted as being the unknown speaker. 44

Figure 18. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JOE. Reference samples from data set 5, test samples from data set 6. 46

Figure 19. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker RHF. Reference samples from data set 5, test samples from data set 6. 47

PAGE

Figure 20. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker HAN. Reference samples from data set 5, test samples from data set 6. 48

Figure 21. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker MOM. Reference samples from data set 5, test samples from data set 6. 49

Figure 22. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker DJB. Reference samples from data set 5, test samples from data set 6. 50

Figure 23. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker MAE. Reference samples from data set 5, test samples from data set 6. 51

Figure 24. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker EHH. Reference samples from data set 5, test samples from data set 6. 52

Figure 25. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker MBB. Reference samples from data set 5, test samples from data set 6. 53

Figure 26. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker LLP. Reference samples from data set 5, test samples from data set 6. 54

Figure 27. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JDM. Reference samples from data set 5, test samples from data set 6. 55

PAGE

Figure 28. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JAC. Reference samples from data set 5, test samples from data set 6. 56

Figure 29. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker BFH. Reference samples from data set 5, test samples from data set 6. 57

Figure 30. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker HS. Reference samples from data set 5, test samples from data set 6. 58

Figure 31. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker BTO. Reference samples from data set 5, test samples from data set 6. 59

Figure 32. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker NAT. Reference samples from data set 5, test samples from data set 6. 60

Figure 33. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker ECJ. Reference samples from data set 5, test samples from data set 6. 61

Figure 34. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker CMW. Reference samples from data set 5, test samples from data set 6. 62

Figure 35. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker SAD. Reference samples from data set 5, test samples from data set 6. 63

PAGE

Figure 36. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JME. Reference samples from data set 5, test samples from data set 6. 64

Figure 37. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JC. Reference samples from data set 5, test samples from data set 6. 65

Figure 38. Confusion matrix for speaker identification experiment with two test speakers, where there was a two-week time interval between the reference samples and the test samples. Reference samples from data set 5 and test samples from data set 2. 70

Figure 39. Plot of sequential analysis results for speaker identification without vowel recognition, and a two-week separation between test samples and reference data. Reference samples from data set 5 and JDM test samples from data set 2. 71

Figure 40. Plot of sequential analysis results for speaker identification without vowel recognition, and a two-week separation between test samples and reference data. Reference samples from data set 5 and BFH test samples from data set 2. 72

LIST OF TABLES

PAGE

Table 1. Data set 1: Breakdown of vowel samples from each speaker in the first 20 recording sessions. 9

Table 2. Data set 2: Breakdown of vowel samples for two speakers recorded after a 2-week interval. 10

Table 3. Correspondence between 2-character vowel codes and phonemic symbols. 10

Table 4. Data set 3: Vowel samples in first half of data base (for reference data). 13

Table 5. Data set 4: Vowel samples in second half of data base (test data). 14

Table 6. Vowel ranking according to two different decision methods. 24

Table 7. Data set 5: Number of vowels in first 1.5 minutes of data base (reference data). 39

Table 8. Data set 6: Number of vowels in second 1.5 minutes of data base (test data). 40

Table 9. Results of text-independent speaker identification without vowel recognition, using a sequential analysis decision procedure. 67

Table 10. Results of text-independent speaker identification without vowel recognition, using a sequential analysis decision procedure. 70

EVALUATION

A method for text-independent speaker identification has been developed utilizing vowel sounds as the basis for recognition. A sequential analysis procedure allows for the accumulation of vowel samples until a particular confidence level (or reject level) is reached. Experimentation using 20 different speakers reached a 95% confidence level of recognition between 5.5 seconds and 90 seconds for 90% of the speakers. The sequential analysis procedure developed in this report provides a practical method for implementing speaker identification.

Robert A. Curtis

ROBERT A. CURTIS, Captain, USAF
Project Engineer

SECTION 1

INTRODUCTION

This is a report on the third in a series of studies on speaker identification based upon a method whereby identification features are extracted from vowel sounds. This particular approach has been the subject of considerable investigation, with two of the more comprehensive studies being those of Paul, et al (1974) and the Aerospace Corp. (1977). The analytical studies reported by Paul resulted in highly successful laboratory experiments and the development of a semi-automatic speaker identification system. However, the Aerospace report noted that there was considerable performance degradation when the system was tested in the field. It was determined that problems such as operator inconsistencies in locating vowel segments, variability in the selection of vowel steady-states, and channel variations all contributed to increasing the error rate. The research described in this report has attempted to address these issues either directly or indirectly through the implementation of additional levels of automatic processing and the application of new concepts in the basic methodology and in the decision process.

One of the primary objectives of this study was to define methods and procedures for performing text-independent speaker identification. Since the basic approach employed the use of vowel sounds, it implied the ability to perform speaker-independent vowel recognition from unconstrained speech. This task is usually reserved for a human operator since all the problems of automatic vowel recognition in unconstrained speech have not yet been resolved.

The methods and procedures described in this report were tested on an experimental data base of vowel samples taken from conversational speech recordings. Five vowel classes were represented in the data, these being among the most frequently occurring vowels in spoken English. Data were gathered from 20 speakers (10 male and 10 female) in 20 recording sessions, and from two of the speakers (one male and one female) two weeks later. The resulting data base contains 4786 vowel samples which are thoroughly documented as to attributes such as stress, context, and word in which the vowel occurred.

In this study, the collection of vowel samples was augmented by two automatic procedures which significantly reduced the variability and inconsistencies in vowel labelling. The first procedure which reduced variability in vowel labelling was an automatic boundary detection algorithm. This is essentially a segmentation algorithm which was beneficial in helping the operator locate the desired vowel sounds. While the operator still had to identify the vowels, the algorithm provided a consistent criterion for locating vowel boundaries, based upon the location of maximum spectral change in the speech signal. The boundary detection algorithm is speaker-independent and operates reliably on unconstrained speech.

The second procedure which reduced variability in vowel labelling was an automatic algorithm for locating the most stable portion of a vowel sound. This is an important step because it addresses the sampling problem and results in a consistent criterion for where to perform the analysis within the vowel. The steady-state is designated as that location within a vowel segment where there is minimal spectral change in the speech signal. The boundary detection and steady-state detection programs used in this study were developed under AFOSR contract F44620-74-C-0034.

One of the significant findings of this study was that it may be possible to achieve reliable speaker identification without recognition of the actual vowel categories. Instead, it appears that the mere detection of vowel-like sounds may be sufficient. The feasibility of this concept was demonstrated through experimentation with the conversational speech vowel samples. The elimination of vowel recognition from this approach to speaker identification reduces the procedural complexity of the process, and removes an extra decision stage which would have compounded the overall system error rate.

Another significant outcome of this study was the successful application of sequential analysis to the decision process. Sequential analysis takes into account the statistics of speaker classification for a series of vowel samples, and permits decisions to be made with specified levels of confidence. This decision procedure lends itself well to the problem of text-independent speaker identification, especially when vowel recognition is not required. As a result, the identification of an unknown recording can proceed by testing each vowel sound as it is detected and making a decision as soon as the desired level of confidence is reached. This approach is time sequential and has a variable time to decision, thereby permitting decisions to be made quickly on speakers who are distinct. Even more important, it is possible that a modified seouential analysis could be made insensitive to channel variability, which is a critical problem in speaker identification.

The entire speaker identification procedure could be made automatic with the addition of one more process, i.e., automatic vowel detection (without recognition). Such an algorithm was beyond the scope of this study, but is not an unreasonable undertaking for future studies.

The next four sections of this report describe the technical accomplishments and the results of the study. Section 2 describes the methods and procedures used, including the methodologies used in locating and labelling the vowel samples, the analysis parameters, and the distance measure. Section 3 contains the experimental procedures and results which led to the conclusion that vowel recognition could be eliminated. Section 4 describes the sequential analysis decision procedure and presents experimental results which demonstrate its application to text-independent speaker identification without vowel recognition. A summary of the work and recommendations for further study are contained in Section 5.

SECTION 2

METHODS AND PROCEDURES

2.1 Experimental Data Base

The speech material consists of recordings of conversational interviews with 20 subjects, ten male and ten female. The interview sessions were informal, with the interviewer asking some general questions and the subject responding in a fashion that usually lead to a conversational situation. The data is a representative sampling of casual speech that would be found in natural language communication. There was no control over the content of the speech material other than the general semantic direction determined by the question. Two of the subjects (one male and one female) were interviewed again two weeks later so that some data reflecting changes over time would be available.

While each interview lasted at least 12 minutes, a 3-minute segment from the initial portion of each was used as a source for vowel samples. A 1.5-minute segment was used from the two repeated interviews. These 22 segments provided a total of 63 minutes of recorded conversation, consisting mostly of subject speech, plus some instances of extended pauses, laughter, and interviewer questions or comments.

2.1.1 Labelling procedure

Each of the 22 segments was digitized at a sampling frequency of 10 kHz and stored in computer files. Using the Interactive Laboratory System (ILS) described by Pfeifer (1977), the speech files were scanned for occurrences of the vowels /i,ɪ,ɛ,æ,ə/, which were then labelled using the following semiautomatic procedure:

1. Using computer playback of the speech, locate a word containing a vowel or vowels of interest.

2. Display the acoustic wave of the word on a graphics display terminal.

3. Execute an automatic algorithm which marks potential boundaries between sounds based upon locations of maximum spectral change in the signal.

4. Verify and label those segments which correspond to the vowels of interest.

The particular set of vowels in the current data base were chosen for labelling because they are among the most frequently occurring in conversational speech and would result in a maximum of vowel samples per class for a given amount of speech. There is one label for each vowel sample.

2.1.2 Label format

A label consists of two lines of text (ASCII) and it contains two basic kinds of information about a speech event, 1) its description, and 2) its location. Figure 1 illustrates the format of a label and what it might contain for the description of a sub-word sound unit. The description of the event is divided into the following six fields which make up the first line of the label.

1. Segment identification (20 characters) - contains some predefined set of symbols which identify the event being labelled. For example, a label for the vowel /i/ would contain the 2-character code /IY/.

2. Stress (two digits) - if the event has some stress level which can be quantified, it can be entered in this field.

3. Environment (eight characters) - identification codes of the events adjacent to the labelled event can be stored. This field is symmetrically defined such that the four leftmost characters are for the left environments and the four rightmost characters are for the right environments.

4. Sequence number (five digits) - labelled events can be numbered if desired, thus providing a numerical indexing of the speech data.

5. Word orthography (20 characters) - a phonemic label will contain the orthographic spelling of the word in which the event occurred.

6. Speaker initials (eight characters) - the initials of the speaker to whom the labelled event belongs.

```
SEGMENT                 ENVIRONMENT         WORD IN WHICH
IDENTIFICATION          OF SEGMENT          SEGMENT OCCURRED
(20 CHARS.)             (8 CHARS.)          (20 CHARS.)
    :                       :                    :
    :                       :                    :
    :                       :                    :
    :        STRESS LEVEL   :   REFERENCE OR     :     SPEAKER
    :        (2 DIGITS)     :   SEQUENCE NO.     :     INITIALS
    :                 :     :   (5 DIGITS)       :     (8 CHARS.)
    :                 :     :       :            :           :
    :                 :     :       :            :           :
    :                 :     :       :            :           :
    :                 :     :       :            :           :
----------------------------------------------------------------
                    ;  ;        ;      ;                   ;
----------------------------------------------------------------
1st line of label


2nd line of label
----------------------------------------------------------------
         ;         ;     ;                         ;         ;
----------------------------------------------------------------
   :        :        :           :                  :          :
   :        :        :           :                  :          :
   :        :        :           :                  :          :
   :        :        :           :                  :          :
STARTING    :    SAMPLING        :                             :
POINT OF    :    FREQUENCY       :           DATE SEGMENT      :
SEGMENT     :    (5 DIGITS)      :           WAS LABELLED      :
(9 DIGITS)  :                    :           (9 CHARS.)        :
            :                    :                             :
            :                    :                             :
     NUMBER SAMPLE          NAME AND LOCATION       INITIALS OF
     POINTS IN SEGMENT      OF FILE                 PERSON DOING
     (9 DIGITS)             (24 CHARS.)             LABELLING
                                                    (8 CHARS.)
```

Figure 1. Format of speech-event label.

The second line of a label contains information pertinent to the location of the event in the digitized speech file, plus some documentary items. There are six fields in the second line.

1. Starting sample point (nine digits) - sample point number of the initial boundary of the event. This number specifies the absolute starting location of the sound in the digitized speech file, thereby providing a time reference of when the sound occurred.

2. Number points (nine digits) - number of sample points in the event. This number indicates the duration of the sound which is labelled.

3. Sampling frequency (five digits) - sampling frequency at which the speech was digitized, for example, 10000 Hz.

4. File name (24 characters) - name of the digitized speech file in which the labelled event is located.

5. Date (nine characters) - date the label was created.

6. User initials (eight characters) - initials of the person doing the labelling.

When labelling a speech event the user is asked to supply the information for the first five fields of the first line of a label. The remaining information in the label is supplied by the computer. Labels provide a convenient method of information retrieval based upon any of the descriptive fields in the first line of the label. The labels describing our conversational speech data are intended to be independent descriptive units so that speech events can be labelled at any level necessary to fulfill the needs of both acoustic-phonetic and linguistic processing. While labelling is usually done at the sub-word level for acoustic-phonetic studies, it is proposed that higher level units such as words, pauses, phrases and sentences could also be labelled so that when studying acoustic measurements, as many parallel levels of description as possible can be provided.

2.1.3 Vowel counts

The five-vowel data base consists of 4786 labelled vowel samples. A breakdown of the number of samples per speaker is given in Table 1 (4538 samples) and Table 2 (248 samples). Table 1 lists the vowels from the first 20 recording sessions, hereafter referred to as data set 1. Table 2 lists the vowels from the speakers recorded two-weeks later, hereafter referred to as data set 2. Each vowel sample has been completely labelled with identification codes, stress, environment, sequence number, word, and speaker initials (except approximately 10 percent which are missing stress). Three levels of stress were specified, with 0 being the lowest and 2 being the highest. A two-character ARPAnet coding scheme was used to represent the vowel sounds and their environmental specifications. Table 3 lists the 2-character codes and their corresponding phonemic symbol. The 2-character codes will be used throughout the remainder of the report.

Once all the vowel segments were labelled, another complete set of labels (one per vowel) was generated which describe the location of the most stable position of the vowel. These steady-state labels were derived automatically by an algorithm which marks the location of least spectral change within the vowel segment.

2.2 Analysis Parameters

A steady-state label represented a single 20 msec frame in the vowel sound. Each vowel steady-state was subjected to the autocorrelation method of linear prediction. There were 200 sampled data points in the analysis window, the data were preemphasized and multiplied by a Hamming window. The output of the analysis was a 12-dimensional vector of reflection coefficients. These coefficients were used as the basic feature vectors for the ensuing speaker identification experiments. It is important to be able to operate successfully with such parameters because they are currently the foundation of low bit-rate vocoder communication channels. Their advantages are that they are bounded between +1 and -1, they exhibit minimal sensitivity to quantization, and they can be tested for instability.

As related to speaker identification, the reflection coefficients represent the spectral properties of the speech signal, and thus they contain speaker-dependent characteristics. This is verified by the fact that perceptual speaker identity is retained in vocoded speech, although there are certainly contributions from other attributes such as fundamental frequency and speaking rate.

Table 1. Data set 1: Breakdown of vowel samples from each speaker in the first 20 recording sessions.

	VOWELS					
SPEAKER	IY	IH	EH	AE	AX	TOTAL
JOE	38	49	35	18	50	190
RHF	84	112	58	34	90	378
HAN	59	61	54	43	49	266
MOM	36	67	43	42	55	243
DJB	37	53	42	32	57	221
MAE	50	51	54	25	46	226
EHH	40	41	34	31	52	198
MBB	59	56	38	21	52	226
LLP	35	58	43	35	54	225
JDM	41	57	30	41	89	258
JAC	51	57	46	24	48	226
BFH	50	75	38	61	69	293
HS	46	91	60	27	66	290
BTO	31	54	42	20	56	203
NAT	34	52	45	23	35	189
ECJ	46	47	55	38	49	235
CMW	30	61	21	18	35	165
SAD	27	23	32	15	36	133
JME	33	33	26	28	41	161
JC	58	56	26	33	39	212
TOTALS	885	1154	822	609	1068	4538

Table 2. Data set 2: Breakdown of vowel samples for two speakers recorded after a 2-week interval.

	VOWELS					
SPEAKER	IY	IH	EH	AE	AX	TOTAL
JDM	25	40	13	17	38	133
BFH	24	29	12	11	39	115
TOTALS	49	69	25	28	77	248

Table 3. Correspondence between 2-character vowel codes and phonemic symbols.

2-character code	phonemic code
IY	i
IH	ɪ
EH	ɛ
AE	æ
AX	ə

2.3 Distance Measure

Each vowel sample was represented by a 12-dimensional vector of reflection coefficients, upon which all future processing was performed. Subsets of the samples were designated for training data and separate subsets were designated as test data. The distance metric used in this study was the weighted Euclidean distance. With this form of distance measure, the reference data for each speaker is represented by a mean vector (as with the unweighted Euclidean distance) and an inverse covariance matrix. The inverse covariance matrix is symmetrical, therefore only half of it need be stored. The distance between reference speaker j and the unknown (or test) vector y is given by

$$D_j = [(y-X_j)^T W_j^{-1} (y-X_j)]^{1/2} \qquad (1)$$

where

X_j = mean reference vector for speaker j

y = test or unknown vector

W_j^{-1} = pooled intra-speaker inverse covariance matrix

$(\ )^T$ = vector transpose.

The matrix W_j is defined as

$$W_j = 1/N_j \sum_{i=1}^{N_j} [x_{ij}-X_j][x_{ij}-X_j]^T \qquad j=1,2,\ldots,k \qquad (2)$$

where X_j is

$$X_j = 1/N_j \sum_{i=1}^{N_j} x_{ij} \qquad j=1,2,\ldots,k \qquad (3)$$

and where k is the number of classes, N_j is the number of vector samples from class j, and x_{ij} is the <u>ith</u> vector sample from speaker j.

In a closed-choice speaker identification task, where the test sample is assumed to be a member of the set of reference samples, a minimum distance criterion is sufficient for the classification process.

2.4 Separation of Test and Reference Data

Prior to running any speaker identification experiments the data base of vowel samples was split into two independent sets. Data set 3 was used for design (from which reference patterns would be derived) and data set 4 was used for testing. Splitting the data base was done on an individual vowel basis for each speaker. If there were N samples and N was even then each set received N/2 samples. If N was odd, then data set 3 received the first (N-1)/2+1 samples. The only constraint on separating the samples was that in data set 3 there be at least 13 samples of each vowel from each speaker. This was a requirement for computing an inverse covariance matrix from 12-dimensional vectors (12 reflection coefficients). Thus if N was less than 25, the first 13 samples were placed in data set 3, and the remaining N-13 samples placed in data set 4. All of the samples from the two-week later interviews (data set 2, Table 2) were used as test data.

A total of 2317 vowel samples were placed in data set 3. A breakdown of vowel counts for data set 3 is given in Table 4. On the average, there were 115 total vowel samples per speaker, with the vowel /AE/ having the least (averaging 16.5 samples per speaker) and the vowel /IH/ having the most (averaging 29.25 per speaker). A total of 2221 vowel samples were placed in data set 4. A breakdown of vowel counts for data set 4 is given in Table 5.

Table 4. Data set 3: Vowel samples in first half of data base (for reference data).

	VOWELS					
SPEAKER	IY	IH	EH	AE	AX	TOTAL
JOE	19	25	18	13	25	100
RHF	42	56	29	17	45	189
HAN	30	31	27	22	25	135
MOM	18	34	22	21	28	123
DJB	19	27	21	16	29	112
MAE	25	26	27	13	23	114
EHH	20	21	17	16	26	100
MBB	30	28	19	13	26	116
LLP	18	29	22	18	27	114
JDM	21	29	15	21	45	131
JAC	26	29	23	13	24	115
BFH	25	38	19	31	35	148
HS	23	46	30	14	33	146
BTO	16	27	21	13	28	105
NAT	17	26	23	13	18	97
ECJ	23	24	28	19	25	119
CMW	15	31	13	13	18	90
SAD	14	13	16	13	18	74
JME	17	17	13	14	21	82
JC	29	28	13	17	20	107
TOTALS	447	585	416	330	539	2317

Table 5. Data set 4: Vowel samples in second half of data base (test data).

	VOWELS					
SPEAKER	IY	IH	EH	AE	AX	TOTAL
JOE	19	24	17	5	25	90
RHF	42	56	29	17	45	189
HAN	29	30	27	21	24	131
MOM	18	33	21	21	27	120
DJB	18	26	21	16	28	109
MAE	25	25	27	12	23	112
EHH	20	20	17	15	26	98
MBB	29	28	19	8	26	110
LLP	17	29	21	17	27	111
JDM	20	28	15	20	44	127
JAC	25	28	23	11	24	111
BFH	25	37	19	30	34	145
HS	23	45	30	13	33	144
BTO	15	27	21	7	28	98
NAT	17	26	22	10	17	92
ECJ	23	23	27	19	24	116
CMW	15	30	8	5	17	75
SAD	13	10	16	2	18	59
JME	16	16	13	14	20	79
JC	29	28	13	16	19	105
TOTALS	438	569	406	279	529	2221

SECTION 3

DEVELOPMENT OF TEXT-INDEPENDENT SPEAKER IDENTIFICATION WITHOUT VOWEL RECOGNITION

3.1 Confusion Matrix Interpretation

Any one row of a confusion matrix can be thought of as a form of histogram, where there is a bin for each column or reference. Figure 2(a) illustrates an example of a confusion matrix which has two rows (test classes), six columns (reference classes), and some hypothetical results of a minimum distance decision process. Since each cell of the matrix is a frequency of occurrence counter, the rows of the matrix can be transformed into the two histograms shown in Figure 2(b). In this example, when 28 samples from class B are tested, 15 are classified as belonging to class B and the remaining 13 samples are scattered among other reference classes. By traditional scoring, 53 percent correct classification might be considered unacceptably low. However, it is not the classification of any one input sample that is significant, for it may not be wise to make an identification decision based upon a single vowel sample. Instead, it is better to examine the accumulated results from the classification of many vowel samples and then make an identification decision based upon the distribution of the individual classifications. This is an acceptable concept for it means that the decision would be global to a particular amount of speech material. Thus, the test data from speaker B in Figure 2 might consist of 28 vowels from 20 seconds of speech, and based upon a simple majority decision or the mode of the classifiation counts, it could be concluded that the 20 seconds of speech belongs to speaker B. The same conclusion could be reached with the 25 test samples from speaker D.

3.2 Experiments Assuming Vowel Recognition

The first experiment was performed under the assumption that vowel recognition could be performed. Each of the 20 speakers was represented by five references, one for each of the five vowels. If a test vowel from an unknown speaker could first be categorized as to vowel class, then it would be matched against the 20 references (one for each speaker) for the corresponding vowel class. Thus, occurrences of the vowel /IY/ would only be matched against references for the vowel /IY/, etc. To simulate a vowel classification process, each vowel category was tested separately, thereby giving an indication of its utiltiy in classifying speakers. The results of the

REFERENCE

TEST	A	B	C	D	E	F		
B	4	15	1	6		2	28	15/28=53.5%
D	2		3	20			25	20/25=80.0%

(a)

(b)

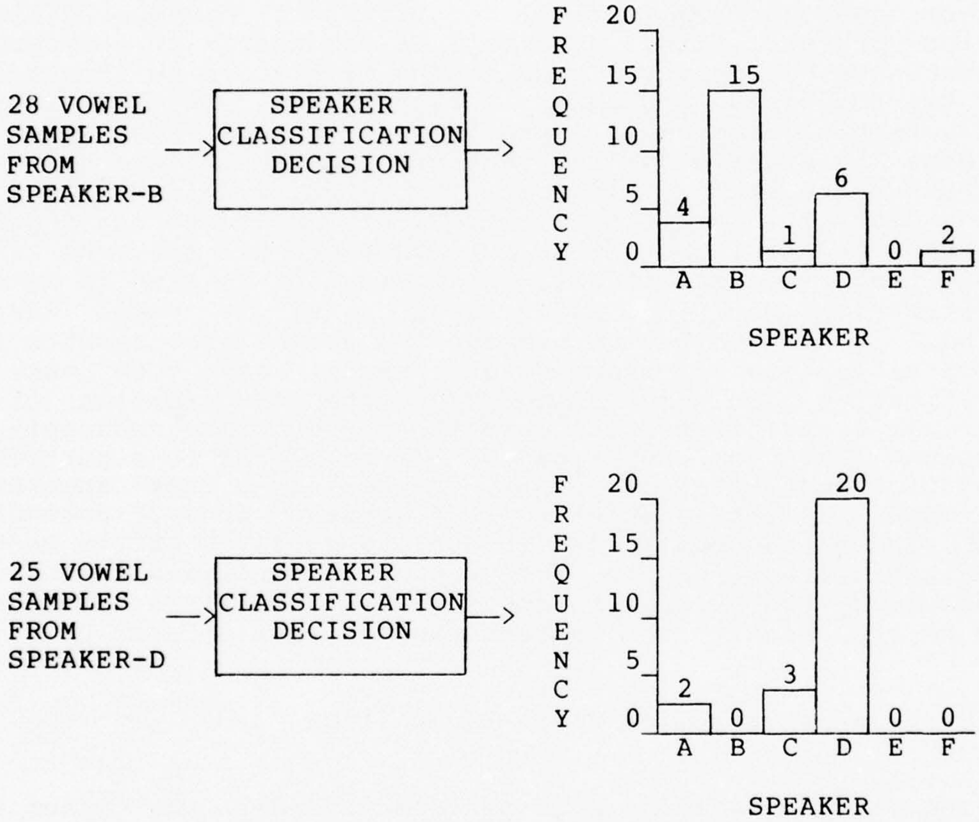


Figure 2. (a) Hypothetical confusion matrix results of a speaker identification experiment having two test speakers and five references.
(b) Histogram interpretation of the results for each test speaker, with each reference being a bin in the histogram.

individual vowel category tests were then grouped to provide overall 5-vowel results based upon the assumption of correct vowel recognition. A minumum distance criterion was used to classifiy each input vowel sample as one of the 20 speakers.

3.2.1 Results using vowel /IY/

When the 438 samples of the vowel /IY/ in data set 4 were compared with the 20 references for the vowel /IY/, 169 were correctly classified (38.58 percent). The confusion matrix for this test condition is shown in Figure 3. The mode of each row of the matrix has been indicated by putting a box around the element with the most classifications. If the final identification decision for each test speaker is based upon the mode of the samples from that speaker, then it can be seen that 11 of the modes fall on the diagonal of the matrix, or that 11 of the 20 input speakers were correctly identified. In the case of speaker JOE, there is an ambiguous decision because of a bimodal distribution, resulting in a tie.

3.2.2 Results using vowel /IH/

When the 569 test samples of the vowel /IH/ were compared with the 20 /IH/ references, 205 were correctly classified (36.03 percent). The confusion matrix for the /IH/ experiment is given in Figure 4. When identification decisions are based upon modal analysis, 16 of the 20 input speakers are correctly identified. There were three tie situations, two of which had the correct speaker in the pair. As compared to the results of the previous vowel test, /IH/ actually had a lower individual vowel classification score than /IY/, but yet in terms of the modal distribution rule, /IH/ gave a significantly higher identification score.

3.2.3 Results using vowel /EH/

Out of the 406 test samples of the vowel /EH/, 158 were correctly classified (38.92 percent). The confusion matrix for the /EH/ experiment is given in Figure 5. Using modal analysis, only 12 of the 20 test speakers would be correctly identified. This vowel had a higher classification score than the vowel /IH/, based upon individual vowel samples, yet the modal decision rule resulted in fewer correct identifications.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	5		1			1		3			1			1		5				2
RHF		16	3		4	11	5	1		1						1				
HAN			25		2	1	1													
MOM			1	1	2	8	2	3	1											
DJB			3		5	3	1	2	1							3				
MAE		2	2		2	10	1	2	2	1						2				1
EHH		1	2			6	4	4			1					1				1
MBB			1		3	17	1	5								2				
LLP			5		3	4										2				3
JDM		1	3		1	1	2			11			1							
JAC			1					1			22					1				
BFH			1			1		1				6	9			1				6
HS	1		1		1							2	13							5
BTO	1		1			1			1					4		1				6
NAT			1			1		1					1	1	4	2				6
ECJ			1		1	3	1	1	1		1	3	1	1		9				
CMW	1				1	4			1					1	4	1			2	
SAD	1														4			3	2	3
JME						1					1		1	1	2	1			8	1
JC	2		1		3	1			1							3				18

Figure 3. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /IY/ only. Data set 3 used for reference and data set 4 for test.
CLASSIFICATION SCORE: 169 out of 438 correct (38.58%)
MODAL SCORE: 11 out of 20 correct (55%)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	5	1	1	1				2	4				1	2	3	1	1			2
RHF		25	5	3	4	3		3	1	8			1		1	2				
HAN			14	1	2			1	3	8				1						
MOM		2		15	8			2	3	1				1		1				
DJB			2	1	9	6				2			1		4		1			
MAE		1	1	4	3	4		1	3	2		1			3	2				
EHH		3	1		4	2	1	4	3	1					1					
MBB		3		5		2		6	1	1					3	6				1
LLP			4	3	1				8	1				3	4	2	1			2
JDM			3	3	2					20										
JAC		2		4				2			12			1	2	2	1			2
BFH			1		1				5	3		7	3	3	6	4				4
HS		2								4		1	33	1	3					1
BTO												1		4	10	6	1			5
NAT									2					4	10	2	4			4
ECJ	1		1		1	1			2				2		4	8	1			2
CMW			1						1						11	3	13			1
SAD														1	4	2	3			
JME						1			2	1			1		6	1	1		3	
JC	1	1		1				2					1		7	3	4			8

Figure 4. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /IH/ only. Data set 3 used for reference and data set 4 for test.
CLASSIFICATION SCORE: 205 out of 569 correct (36.03%)
MODAL SCORE: 16 out of 20 correct (80%)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	11								1							1		4		
RHF		11		1	2	6	1		2							5		1		
HAN		2	13		1	2			7						2					
MOM	6	4		6		1			2									2		
DJB			4		5	6			2	1		1				1		1		
MAE		1	4		1	12			7											2
EHH		1	3	1	2	6										3		1		
MBB		1				8		1	3							3		3		
LLP	1		4		1	4			10			1								
JDM		1	4	1		3			2	1								3		
JAC									1		18					1		3		
BFH						2			1		1	5	1		5	2		1		1
HS						2							18	1	2	2		5		
BTO	1							1	1				2	5	7	1		3		
NAT						4			2				3	1	10		1			1
ECJ		1				6						1		1		18				
CMW															1	1		5		1
SAD			1												1	1		13		
JME																1		11		1
JC	4		1			3								1		1		2		1

Figure 5. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /EH/ only. Data set 3 used for reference and data set 4 for test.
CLASSIFICATION SCORE: 158 out of 406 correct (38.92%)
MODAL SCORE: 12 out of 20 correct (60%)

3.2.4 Results using vowel /AE/

The experiment using only the vowel /AE/ resulted in 104 correct classifications out of 279 attempts (37.28 percent). The confusion matrix for the /AE/ experiment is given in Figure 6. This time only nine correct identification decisions would be made using the modal rule. This vowel had a higher score for individual sample classifications than the vowel /IH/, but a much lower identification score based on the modal rule.

3.2.5 Results using vowel /AX/

Out of the 529 test samples of the vowel /AX/, 235 were classified as the correct speaker (44.42 percent). The confusion matrix for the /AX/ experiment is given in Figure 7. This vowel had not only the highest individual sample classification score, but also the highest identification score based on a modal decision rule.

3.2.6 Comparison of vowel results

An overall comparison of the results obtained for each of the five vowel sounds leads to different conclusions, depending upon the method of decision making. Table 6 illustrates how the ranking of the vowels is affected by whether the ranking is based upon the classification of individual samples or the modal decision rule. The vowel /AX/ is the only one which maintains its ranking position in both cases, while /EH/, /IY/, and /AE/ maintain their rankings relative to each other. /IH/ is the only vowel which undergoes a dramatic shift. The actual ranking of the vowels is not an issue at this time, however. Instead it is more important to note the superiority of results based upon the modal decision rule.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE				2			2		1											
RHF		2	1	1	2		4		1	3						3				
HAN			9	1	1		3		3	2		1				1				
MOM			2	8			1			2						5				3
DJB		1		1	3	1	2		2	2		1								3
MAE		1					3		2	2		2				2				
EHH							7		3			1				4				
MBB				1			4		1							2				
LLP		1	2						11	3										
JDM										19						1				
JAC				1			1				2	1				4				2
BFH			1				1					21				6			1	
HS			1									8				3			1	
BTO							2		1	1		1		1		1				
NAT		1	1						1			2				3	1			1
ECJ		1	3				1					2				12				
CMW												1				1	1	1		1
SAD									1											1
JME										1		8				1	1		3	
JC		1			1		2					3				4				5

Figure 6. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /AE/ only. Data set 3 used for reference and data set 4 for test.
CLASSIFICATION SCORE: 104 out of 279 correct (37.28%)
MODAL SCORE: 10 out of 20 correct (50%)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	1		4	1		1		6	7	1			2				1		1	
RHF	1	19	1		2	5	1	1		7		1	2	1	2	2				
HAN	1		10		3	3		1	2	2			1			1				
MOM	1	2	1	11	1			1	1	3	1		3			2				
DJB			4		11	3	2		1	5						2				
MAE			2		3	6		4	2	2		1	1			2				
EHH	1	2	3	2	3	1	7	1		4						2				
MBB	2	1		3	2	3		8		1		1			2	1			2	
LLP	1		5		2	4			7	1			3	2		1	1			
JDM		1	1	1	1		1	2	1	34		1							1	
JAC		1		2	1						18	1	1							
BFH		1			1				3			26	2			1				
HS		2	2		1				1	1		7	18	1						
BTO	1	1						1	2			2	1	14	4		2			
NAT		1				1				1		1	2	6	2		1		1	1
ECJ		2	1			3				2			1			15				
CMW										2				4			11			
SAD										1		1	2	3			5	5	1	
JME		2				1				1		2		1			3		10	
JC	1				1	8		1		2		3	1							2

Figure 7. Confusion matrix for speaker identification assuming vowel recognition. These results are for the vowel /AX/ only. Data set 3 used for reference and data set 4 for test.
CLASSIFICATION SCORE: 235 out of 529 correct (44.42%)
MODAL SCORE: 17 out of 20 correct (85%)

Table 6. Vowel ranking according to two different decision methods.

CLASSIFICATION OF EACH SAMPLE		MODAL DECISION RULE	
VOWEL	% CORRECT	VOWEL	% CORRECT
AX	44.42	AX	85
EH	38.92	IH	80
IY	38.58	EH	60
AE	37.28	IY	55
IH	36.03	AE	45

3.2.7 Combined vowel results

The results of the separate vowel experiments can be combined to demonstrate the effect of using all the vowels in making an identification decision. There was a total of 2221 test samples from all five vowel categories. On the basis of classifying each vowel sample according to speaker, the overall score can simply be computed from the known scores of the separate vowel experiments, i.e., 871 correct out of 2221 (39.2 percent). If the modal decision rule is applied to the combination of all five vowels, the results are significantly improved. A composite confusion matrix for all five vowels is shown in Figure 8. This matrix is simply the sum of each of the individual vowel matrices. The mode of each row of the matrix has been outlined, and it can be seen that a modal decision rule would make only one error out of 20 decisions.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	22	1	6	4		2	2	11	13	1	1		3	3	3	7	2	4	1	4
RHF	1	73	10	5	14	25	11	5	4	19		1	3	1	3	13		1		
HAN	1	2	71	2	9	6	4	2	15	12		1	1		2	2				
MOM	7	8	4	41	11	9	3	6	7	6	1		3	1		8		2		3
DJB		1	13	2	33	19	5	2	6	10		2	1		4	6	1	1		3
MAE		5	9	4	9	32	4	7	16	7		4	1		3	8				3
EHH	1	7	9	3	9	15	19	9	6	5	1	1			1	10		1		1
MBB	2	5	1	9	5	30	5	20	5	2		1			5	14		3	2	1
LLP	2	1	20	3	7	12			36	5		1	3	5	4	5	2			5
JDM		3	11	5	4	4	3	2	3	85		1	1			1		3	1	
JAC		3	1	7	1		1	3	1		72	2	1	1	2	8	1	3		4
BFH		1	3		2	3	1	1	9	3	1	65	15		11	14		1	1	11
HS	1	4	4		2	2			1	5		18	82	3	5	5		5	1	6
BTO	3	1	1			1	2	2	5	1		4	3	28	21	9	3	3		11
NAT		2	2			6			5	1		3	6	12	26	7	7		1	13
ECJ	1	4	4		2	13	2	1	3	2	1	6	4	2	4	53	1			2
CMW	1		1		1	4			2	2		1		5	16	6	25	6	2	3
SAD	1		1						1	1		1	2	4	9	3	8	21	3	4
JME		2				3			2	3	1	10	2	2	8	4	5	11	24	2
JC	8	2	2	1	5	12	2	3	1	2		6	2	1	7	11	4	2		34

Figure 8. Composite confusion matrix for speaker identification test assuming recognition of five vowel categories. There were five references for each speaker, one corresponding to each vowel class. Reference data taken from data set 3, test data from data set 4. Each input test sample was first classified according to vowel category and then matched with those references for that vowel only.
CLASSIFICATION SCORE: 862 out of 2221 correct (38.8%)
MODAL SCORE: 19 out of 20 correct (95%)

3.2.8 Statistical stability of results

At this point some comments are in order regarding the interpretation of the results. In data set 3 there were several sets of vowel samples which contained the minimum number of samples to invert the covariance matrix. For example, the vowel /AE/ had the least number of samples, and therefore the results based upon this vowel may be statistically unreliable. The stability of the results were partially checked by using the vowel samples from the recordings made two weeks later (data set 2) as test data and matching against two different groups of references. Reference group A was made up from data set 3, the same as used in the experiments thus far described. Reference group B was made up of all the vowel samples on both data set 3 and data set 4. Thus reference group B was computed from approximately twice as many vowel samples as reference group A. The test data consisted of vowel samples from one male and one female speaker. The experiment assumed vowel recognition, therefore each vowel category was tested separately and the results of all five vowel tests combined to indicate overall performance.

Figure 9 contains the confusion matrix for each vowel test and the combined results for speaker JDM when compared with reference group A. In this case 66.92 percent of the test vowels were correctly classified. According to a modal decision rule the correct speaker would be chosen in three of the five vowel categories. When the results from individual vowel categories are combined, the mode of the classifications still remains with the true speaker. Figure 10 illustrates the results of using the exact same test data from speaker JDM, but comparing it with reference group B. This time 65.41 percent of the individual samples were correctly classified, only a 1.5 percent change from the previous experiment. However, the mode of the classifcations in each vowel category is located at the true speaker. The mode for the combined results is still located with the true speaker, but it is interesting to note that the number of classifications for the next nearest speaker has dropped from 25 with reference group A to 16 with reference group B.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
IY			14		2	2				6										1
IH			2		1					37										
EH			8			3			1				1							
AE			1							16										
AX		1			3	2			1	30			1							
JDM		1	25		6	7			2	89			2							1

Figure 9. Confusion matrix for speaker identification experiment using JDM test data (data set 2), and approximately 1.5 minutes of reference data (data set 3) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results.
SCORE: 89 correct out of 133 (66.92%)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
IY		1	6			2			1	14			1							
IH			1	1	10		1			27										
EH					1	4				8										
AE		1			2	1			1	12										
AX		1			3	3			1	26			1			3				
JDM		3	7	1	16	10	1		3	87			2			3				

Figure 10. Confusion matrix for speaker identification experiment using JDM test data (data set 2), and approximately 3.0 minutes of reference data (data set 1) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results.
SCORE: 87 correct out of 133 (65.41%)

The results of comparing the BFH test samples (from 2nd session, 2 weeks later) with reference group A are shown in the confusion matrix in Figure 11. Three of the five vowels had the mode of classifications with the true speaker and the combined score is 38.26 percent correct. The mode of the classifications for the combined results is also located at the true speaker. When the same BFH test samples were compared with reference group B, three vowels had the mode of classifications located at the true speaker (see Figure 12). These are the same three vowels as the reference group A experiment. The combined result of 38.26 percent correct is also identical to the combined result from the reference group A experiment. However, the speaker with the next highest number of classifications went from a count of 18 with reference group A to a count of 26 with reference group B. This trend is the reverse of that found for speaker JDM. The results of the experiments with reference groups A and B have shown that for the male and female speakers tested, there was essentially no change in either classification scores or the mode of the classifications when the number of samples in the reference was approximately doubled.

3.2.9 Effects of time

The experiments with reference groups A and B also provide information on how well this approach to speaker identification might perform over time, since the test data in the experiments was recorded two weeks after the data used in making up the references. Reference group A was computed from data set 3, therefore the results of using data set 3 for reference and data set 4 for test can be directly compared with those using reference group A. The confusion matrix for the two speakers JDM and BFH from data set 4 is given in Figure 13(a). This matrix is taken from two rows of the confusion matrix of Figure 8. The confusion matrix for speakers JDM and BFH two weeks later is shown in Figure 13(b). Each matrix represents the combined results of all five vowel categories, assuming vowel recognition. When the test data is from the same recording session (Figure 13(a)) 85 out of 127 samples were correctly classified (66.93 percent) for the JDM test data and 65 out of 145 samples were correctly classified (44.23 percent) for the BFH test data, giving an overall score of 57.69 percent correct. For both test speakers, the mode of the correct classifications is located at the true speaker. When the test data is from recording sessions two weeks later (Figure 13(b)), 89 out of 133 samples were correctly classified (66.92 percent) for speaker JDM and 44 out of 115 were correctly classified (38.26 percent) for speaker BFH, giving an overall score of 53.63 percent correct.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
IY					3							3	5			6				7
IH						1			1	1		7	6	1	5	4	1			2
EH						2						1			3	2		2		2
AE												8				3				
AX						1		2	1	1		25	3	1		3	1	1		
BFH					3	4		2	2	2		44	14	2	8	18	2	3		11

Figure 11. Confusion matrix for speaker identification experiment using BFH test data (data set 2), and approximately 1.5 minutes of reference data (data set 3) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results.

SCORE: 44 correct out of 115 (38.26%)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
IY					3				1			7				9				4
IH					1				4			13	3		2	5	1			
EH												2		1	1	3		2		3
AE					1			1				4				4				1
AX		2	1			1		2	6			18	2	2		5				
BFH		2	1		5	1		3	11			44	5	3	3	26	1	2		8

Figure 12. Confusion matrix for speaker identification experiment using BFH test data (data set 2), and approximately 3.0 minutes of reference data (data set 1) from 20 speakers. The first five rows of the matrix give speaker classification results for the separate vowels, while the last row indicates the combined vowel results.

SCORE: 44 correct out of 115 (38.26%)

(a)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JDM		3	11	5	4	4	3	2	3	85		1	1			1		3	1	
BFH		1	3		2	3	1	1	9	3	1	65	15		11	14		1	1	11

(b)

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JDM		1	25		6	7			2	89			2							1
BFH					3	4		2	2	2		44	14	2	8	18	2	3		11

Figure 13. Comparison of two speaker identification experiments having the same reference set (data set 3) and two test speakers (one male and one female).

(a) Test data from the same recording session as reference data, but different time segments. Test samples from data set 4.

(b) Test data from a recording session two-weeks after the reference data was recorded. Test samples from data set 2.

It is expected that there would be some level of degradation as the time between the reference data and the test samples increases. Speaker JDM appears to be rather well separated from the other speakers in the population and there is virtually no change in the classification scores after a two week period. There is, however, a slight decrease in the strength of the mode of the classifications. With the test data from the same recording session as the reference, the next nearest speaker had a count of only 11 classifications, whereas that count increased to 25 when data from two weeks later were tested.

After a time period of two weeks, the results of individual sample classifications for speaker BFH decreased from 44 percent to 38 percent. The mode of the classifications still remains with the true speaker, however, the strength of that mode has decreased because the next nearest reference had a classification count of 18 as compared to a count of 14 when the test and reference data occurred in the same recording session.

3.2.10 Discussion

The results thus far have been very encouraging, considering the fact that the vowel samples were from conversational speech and there were no text-dependent restrictions placed on the experiments. The fact that truly independent test and reference data were used and that results (for two speakers) did not significantly degrade after a two-week time interval further suggests that this approach will be successful in practical speaker identification situations. One rather severe assumption which was made in the previous experiments, however, was that each vowel sample could be correctly recognized as to vowel category so that it could be compared with the corresponding vowel reference from each speaker. Since we are dealing with conversational speech from a variety of speakers, the need for general-purpose speaker-independent vowel recognition casts some doubt on the possibility of a totally automatic speaker identification system with this approach.

Assuming some form of vowel recognition was available, it is possible that vowel recognition errors could compound speaker identification errors. Even our own data base, with its five vowel categories, is subject to perceptual vowel recognition errors, since the listener's judgement of vowel categories has been imposed on the data.

3.3 Speaker Identification Without Vowel Recognition

3.3.1 Method I - Multiple patterns for each reference speaker

One way of circumventing the recognition of each test vowel sample is to allow each reference speaker to be represented by several reference patterns, *e.g.*, one for each vowel category. Then, without regard for input vowel category, compare each test vowel sample with all references for all speakers. For each test sample the number of distances to be computed increases from M, the number of reference speakers, to MV where V is the number of vowel categories. Thus the computation time increases by a factor of V. However, the vowel recognition is now confined to the formulation of the reference patterns, which makes the procedure more feasible. Since distances to all vowel categories will be computed, the minimum distance criterion will allow an input sample the opportunity to seek the reference pattern representing not only the correct speaker, but also the correct vowel category.

This concept was investigated in an experiment where a reference pattern was computed for each of the five vowel categories from each of 20 speakers, using the vowel samples in data set 3. This gave a total of 100 reference patterns. Each of the samples in data set 4 was compared with the 100 references without any attempt to categorize the test samples. Based upon the minimum distance criterion, 923 out of 2221 samples (41.56 percent) were correctly classified as the true speaker. This is a very interesting result because a higher percentage of the samples were correctly classified when there was no vowel recognition on the test data then when there was vowel recognition (38.8 percent). The confusion matrix for this experiment is given in Figure 14. An examination of the mode of the classifications for each speaker shows that for 18 of the 20 speakers the correct speaker would be chosen by a modal decision rule, which is one less than when the test vowels were recognized. The speaker who was missed by the modal decision rule with vowel recognition was also one of the two missed when there was no vowel recognition.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	27	1	4	2		2		5	16		2		3	7	8	3	4	2		4
RHF	1	81	10	2	17	22	10	3	4	17		2	7		6	6	1			
HAN	2		57	1	14	7		1	9	29			2	1	1		2	3		2
MOM	7	5	1	50	7	4		8	7	15	2	3	2	1	1	5		2		
DJB		2	11	1	32	20	2	1	15	10		3	2	1	3	3	1			2
MAE	1	1	6	2	9	36	2	8	16	10		3	1	3	4	7	1			2
EHH	2	12	9	7	15	8	11	8	3	9	1			1	4	6		1		1
MBB	1	8	3	9	4	31	1	21	3	1		3			7	15			1	2
LLP	1		10	3	8	8			41	2		1	4	6	10	3	2	1	2	9
JDM		2	5	4	5	5	3	2	3	93			3					1	1	
JAC	3	8		3	1			4	1		74	1		3	1	4	2	3		3
BFH		1	3		2	3			10			65	16	6	15	11	1	1	2	9
HS		3	4		4				1	5		11	91	6	4	4		4	1	6
BTO	1	1				1			4	1	1	3	4	37	20	8	7	1		9
NAT		1			1	1		2	2	2		1	4	13	32	9	11	2	1	10
ECJ		2	3		3	5	1	1	2	2	1	4	7	1	3	72	4	1	1	3
CMW			1						3	2			2	4	17	4	34	3	2	3
SAD			3						1	1		1	1	7	7	2	12	22	1	1
JME		5	3			5		1	2	1	3	4	3	3	9	2		16	18	4
JC	6	2	1		2	8		2		1		7	4	4	19	9	7	4		29

Figure 14. Confusion matrix for speaker identification experiment without vowel recognition on the test samples. There were 20 reference speakers, each represented by five patterns (one for each vowel), for a total of 100 reference patterns. Reference samples from data set 3 and test samples from data set 4.
CLASSIFICATION SCORE: 923 out of 2221 correct (41.6%)
MODAL SCORE: 18 out of 20 correct (90%)

3.3.2 Method II - Merging of all vowel samples

While the previous experiment demonstrates the feasibility of speaker identification without having to perform vowel recognition on the test samples, it may still be undesirable to have multiple vowel references for each known speaker. This can be eliminated by representing each speaker with a single reference pattern consisting of the pooled samples of all five vowel categories. Thus, in a practical system, a reference would be developed by simply collecting speech samples of the general category "vowel". Then test data would also consist of samples of the general category "vowel".

An experiment using this approach was carried out using the vowel samples from data set 3 to compute 20 references and the samples from data set 4 for test. Of the 2221 test samples, 999 were correctly classified (44.98 percent). This result is yet another improvement over those requiring total vowel recognition or having vowel recognition in the preparation of speaker references. In addition, these results are statistically sound because of the pooling of the vowel samples. The confusion matrix for this experiment is given in Figure 15. It can be seen that the mode of the classifications is located at the true speaker in all cases except one. This is the same speaker who was missed in both the total vowel recognition experiment and the partial vowel recognition experiment. It has therefore been demonstrated that there is a higher rate of correct speaker classification for individual vowel samples when there is no vowel recognition (44.98 percent) then when vowel recognition is performed (38.8 percent). In terms of using a modal decision rule to actually make the speaker identification decisions, the results are essentially unchanged with or without vowel recognition. But since the number of correct classifications increased when there was no vowel recognition, it is expected that the mode of the classifications for each test speaker would also increase in strength. A comparison of Figure 8 and Figure 15 shows that with no vowel recognition the strength of the mode increased for 15 test speakers, decreased for four speakers, and remained unchanged for one speaker.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	31		1	2	4	3		1	20	1	2			2	2		7	6		8
RHF		43	5	4	23	37	14	7	4	15			2		2	27	2		2	2
HAN	2		46	4	24	10	2		8	26		1				4		4		
MOM	6	3		53	4	9	1	8	8	13	3			1		6	3	1	1	
DJB			5	1	61	16		2	12	3			2	1	1	2				3
MAE		1	5		21	47		5	15	6		1	1			5	1		1	3
EHH	2	2	9	4	19	15	24	1	4	4				1		10	1	1		1
MBB	4	1	2	3	7	32	3	24	4	2	1				2	18		1		6
LLP	2		7	1	8	18			61	1				3	4	3	1			2
JDM			1	2	8	4	5		1	100						4		1		1
JAC	2	3			1	1		2			73			2		13	5	6	3	
BFH					4	2			13			64	7	4	12	23	2	1	4	9
HS		2	3		3	2	2	1	1	1		9	66	6	5	24		12		7
BTO				1					4		1		2	46	16	4	11	3	1	9
NAT						2		1	9			2	6	16	26	3	11	5		11
ECJ		1			2	8	3	2	1			3	2	1	5	84	3			1
CMW														1	7	4	48	12	1	2
SAD	1					1								3		1	17	35	1	
JME							2	8	1		2	1			1	5	9	19	28	3
JC	5		1		1	8		4	5			1	1	1	24	7	3	5		39

Figure 15. Confusion matrix for speaker identification experiment without vowel recognition, but where each reference speaker is represented by the pooled samples of all vowel classes. There is only one pattern for each reference speaker. Reference samples from data set 3 and test samples from data set 4.
CLASSIFICATION SCORE: 999 out of 2221 correct (45%)
MODAL SCORE: 19 out of 20 correct (95%)

3.4 Discussion

The elimination of vowel recognition from the speaker identification process is of great significance. Because the reference for each known speaker consists of pooled samples across vowel categories, each speaker is essentially represented by the long term statistics of vowel steady-states. It is also a way of representing the vowel space of each speaker, which is expected to be strongly speaker-dependent. This explanation cannot be completely supported at this time, however, since the vowel samples used in this speaker identification study represented only five vowel categories. The reference patterns prepared in this particular approach are actually computed from a selected subset of the spectral samples which would be used in the computation of the long-term statistics of the generalized speech signal. Therefore, the amount of variation with which a system must deal is reduced from the total variation in speech to the allophonic variation of vowel steady-states. The individual vowel samples used for testing do not have a statistical basis, but yet they have a strong tendency to be attracted to the vowel space of the true speaker. A statistical approach could be applied by accumulating statistics on many input vowel samples and then using the statistical parameters as test data. This would be one possible way of making a system insensitive to varying channel characteristics. While this approach was not pursued in this study, a sequential decision procedure for the individual vowel samples was investigated and will be presented in the following section.

SECTION 4

SEQUENTIAL ANALYSIS

4.1 Introduction

The simulation of a true text-independent speaker identification situation with conversational speech requires that, for the preparation of speaker references, vowel samples be collected for a certain period of time, and that the test vowel samples be taken from the unknown recording in a sequential fashion as they occur in the speech. As each test sample is classified, the cumulative results of all the classifications can be examined, with the possibility that an identification decision could be made if a sufficient number of samples have been placed with any one reference speaker. This type of sequential analysis procedure was described by Wald (1952) and it is appropriate to this application where there are many samples from one unknown speaker. This is also known as a delayed decision procedure because the actual decision is not made until a sufficient number of individual classifications have been performed.

A major advantage of the sequential analysis procedure is that the number of test samples required to make a decision is variable, and therefore the amount of speech material needed to reach a decision could be small if a speaker is distinct. If, for example, after only 25 test samples, 20 are classified as one particular reference speaker, an identification decision could be made with a high degree of confidence.

4.2 Reorganization of Data Base into Discourse Order

In order to experiment with sequential analysis, the data base had to be organized to represent the true sequence of the vowels as they occurred in the speech recordings, i.e., in discourse order. Data sets 3 and 4 were not adequate for this task because they were formed on the basis of splitting the samples from each vowel category into two equal sets. Therefore the samples were not necessarily in discourse order.

Two new data sets were formed from the total in data set 1, and these were identified as data set 5, which would be used for computation of references, and data set 6, which would be used for testing. This time the split of the samples was done on the basis of time. Since there was a total of three minutes of speech from each speaker, those samples occurring in the first 1.5 minutes of the discourse were placed in data set 5 and those samples occurring in the 2nd 1.5 minutes of the discourse were placed in data set 6. The vowel samples in data set 6 were organized sequentially for each speaker in the order in which they were actually spoken in the recordings. As a result of the split, data set 5 had 2369 vowel samples and set 6 had 2169. Table 7 gives a breakdown of data set 5 by speaker and vowel. Table 8 gives a breakdown of data set 6 by speaker and vowel.

4.3 Method II Results Using Sequential Data

In order to obtain the data for sequential analysis, a complete experiment was performed with all 20 speakers and no vowel recognition. This is equivalent to the experiment with no vowel recognition in section 3.3.2 except that some of the samples which used to be in the test set were shifted to the reference set, and vice versa. Overall, of the 2169 samples tested, 1070 were correctly classified as the true speaker, for a score of 49.33 percent. This score is somewhat higher than when the data was not in discourse order (44.98 percent). The confusion matrix of the individual sample classifications is given in Figure 16. The mode of the classifications for each test speaker has been highlighted, and it can be seen that on a delayed-decision modal rule basis, 19 of the 20 test speakers would be correctly identified. It is interesting that the speaker who was consistently in error on all other experiments is now correctly identified, but an error has occurred with one of the female speakers. The confusion matrix of Figure 16 therefore reflects the results from 1.5 minutes of test data from each speaker and 1.5 minutes of reference data from each speaker. If a decision was made at the end of the 1.5 minutes of testing, there would be 95 percent correct identification.

4.4 Illustration of Sequential Analysis Procedure

There may be occasions when it would be advantageous to make an identification decision sooner than 1.5 minutes, or it might be that there is less than 1.5 minutes of speech material for testing. When such is the case, sequential analysis can be advantageous.

Table 7. Data set 5: Number of vowels in first 1.5 minutes of data base (reference data).

	VOWELS					
SPEAKER	IY	IH	EH	AE	AX	TOTAL
JOE	16	26	19	11	30	102
RHF	45	52	26	19	47	189
HAN	33	29	33	25	23	143
MOM	14	27	19	20	35	115
DJB	17	29	23	17	30	116
MAE	33	32	20	13	23	121
EHH	15	16	20	15	38	104
MBB	31	37	14	10	34	126
LLP	10	29	26	16	20	101
JDM	28	35	18	19	43	143
JAC	26	33	29	14	23	125
BFH	27	35	20	29	38	149
HS	24	43	28	9	35	139
BTO	11	22	21	13	25	92
NAT	19	45	34	13	20	131
ECJ	17	22	27	21	22	109
CMW	14	33	7	11	19	84
SAD	14	13	17	8	21	73
JME	13	22	11	20	27	93
JC	27	30	12	23	22	114
TOTALS	434	610	424	326	575	2369

Table 8. Data set 6: Number of vowels in second 1.5 minutes of data base (test data).

	VOWELS					
SPEAKER	IY	IH	EH	AE	AX	TOTAL
JOE	22	23	16	7	20	88
RHF	39	60	32	15	43	189
HAN	26	32	21	18	26	123
MOM	22	40	24	22	20	128
DJB	20	24	19	15	27	105
MAE	17	19	34	12	23	105
EHH	25	25	14	16	14	94
MBB	28	19	24	11	18	100
LLP	25	29	17	19	34	124
JDM	13	22	12	22	46	115
JAC	25	24	17	10	25	101
BFH	23	40	18	32	31	144
HS	22	48	32	18	31	151
BTO	20	32	21	7	31	98
NAT	15	7	11	10	15	58
ECJ	29	25	28	17	27	126
CMW	16	28	14	7	16	81
SAD	13	10	15	7	15	60
JME	20	11	15	8	14	68
JC	31	26	14	10	17	98
TOTALS	451	544	398	283	493	2169

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JOE	36		4	4	4	2		1	15		1			1		1	4	4	1	10
RHF		50	7	7	28	28	25	4	2	9			3		3	17	1	1	2	2
HAN	2		70		16	5	2		8	10		1			1		4	3		1
MOM	10	3	2	53	7	12	5	6	9	13	1					3		3		1
DJB	1		10		56	12	3	2	13	1		1	1		1	2				2
MAE		1	3		29	44		3	13	2		1	1			5				3
EHH		4	8	5	17	12	22	8	3	5				1		4	1	1	1	2
MBB	1	7	1	1	7	24	1	35	2						2	15		1		3
LLP	2		10		12	13			71					2	3	3	3		1	4
JDM		1	1	3	8	4	1	1	1	91						2		2		
JAC	1	2			1	1		1	1		76				1	7	3	3	2	2
BFH			2		5	4			11			70	9	2	9	19	1	1	3	8
HS		2	5		5		2	1	3	1		8	87	6	5	14	1	5		6
BTO				1	1	1		2	3		2		2	34	41	5	12	3	1	3
NAT								3	5			1	3	4	29	2	2	6		3
ECJ		2	2		2	9	2	1	4		2	3	2	1	6	86	2			2
CMW		1							1					1	11	6	45	13	1	2
SAD												1			1	1	14	41	2	
JME								3	2		2	1			2	5	4	12	35	2
JC	6			1	2	7		4	2			2	2		18	7	4	3	1	39

Figure 16. Confusion matrix for speaker identification experiment without vowel recognition on either the test or reference samples, and the test data presented in discourse order. There were 20 reference patterns, one for each known speaker. Reference samples from data set 5, and test samples from data set 6.
CLASSIFICATION SCORE: 1070 out of 2169 correct (49%)
MODAL SCORE: 19 out of 20 correct (95%)

In order to apply the sequential analysis procedure, we must examine the accumulated intermediate speaker classification results after each test sample has been classified. Thus, the intermediate results reflect the status of the decision process as a function of time. For a given set of test samples from an unknown talker, each test sample is subjected to a minimum distance classification test. Each time a reference speaker is chosen as having the minimum distance, its classification count is incremented by one. The intermediate results can be illustrated graphically by plotting the accumulated number of classifications for each reference speaker as a function of the number of samples tested. What remains then is a decision method for the final identification procedure, based upon the intermediate results.

Wald (1952) developed a criterion whereby a sequential decision threshold can be defined in terms of statistical probabilities. The strength of the decision can be specified according to the desired probability of false rejection, P[FR], and the probability of false acceptance, P[FA]. The classification factor, Q, is defined as

$$Q_j = S_j/m \qquad (j=1,2,...,N)$$

where S_j is the number of classifications for a given reference and m is the number of samples which have been tested. If, for a given reference, $Q_j>P1$ then it might be expected that the unknown speaker is the same as the reference. Alternately, for a given reference, if $Q_j<P0$ then it might be expected that the test and the reference are not the same, so that particular reference should be rejected. If $P1>Q_j>P0$, then no decision is made and the testing is continued. In terms of P[FR], P[FA], P1, and P0, acceptance and rejection thresholds can be established as a function of m, the number of test samples. The acceptance threshold L1 is computed as

$$L1 = \frac{\log\frac{1-P[FA]}{P[FR]}}{\log\frac{P1}{P0} - \log\frac{1-P1}{1-P0}} + m\,\frac{\log\frac{1-P0}{1-P1}}{\log\frac{P1}{P0} - \log\frac{1-P1}{1-P0}}$$

and the rejection threshold L0 is computed as

$$L0 = \frac{\log\frac{P[FA]}{1-P[FR]}}{\log\frac{P1}{P0} - \log\frac{1-P1}{1-P0}} + m \frac{\log\frac{1-P0}{1-P1}}{\log\frac{P1}{P0} - \log\frac{1-P1}{1-P0}}.$$

Thus, after classifying each input sample, the number of classifications S_j, for each reference speaker is compared with L0 and L1. If $S_j>L1$ then there is a strong likelihood that the test speaker is the same as the reference speaker j and the test can be terminated. If $S_j<L0$ then there is a strong likelihood that the test speaker is not reference speaker j. And if $L1>S_j>L0$ then no decision can be made regarding reference j.

The sequential decision procedure is illustrated graphically in Figure 17. The abcissa represents the number of samples tested, and the ordinate represents the classification count (number of minimum distance classifications). The two parallel lines represent the acceptance and rejection decision thresholds. These thresholds were computed according to the following specifications:

$P[FR]=.05$
$P[FA]=.05$
$P1=.35$
$P0=.25.$

Sequential classification results are shown for a hypothetical experiment in which there are 50 vowel samples from an unknown speaker X, and there are three known speakers A, B, and C, and it is known that X is one of the three references. Each of the incrementing lines represents one of the reference speakers, or the number of times out of m attempts that a particular reference was chosen as having the minimum distance. In this example, reference speaker A and reference speaker C were infrequently chosen as the classification choice and therefore they soon crossed the rejection threshold L0. It could therefore be concluded that the unknown speaker was neither A or C, with a .05 probability of false rejection. On the other hand, the input samples were classified as speaker B often enough that the acceptance threshold was crossed after only 35 test samples. Thus the conclusion would be that the unknown speech material came from speaker B, with a .05 probability of false acceptance.

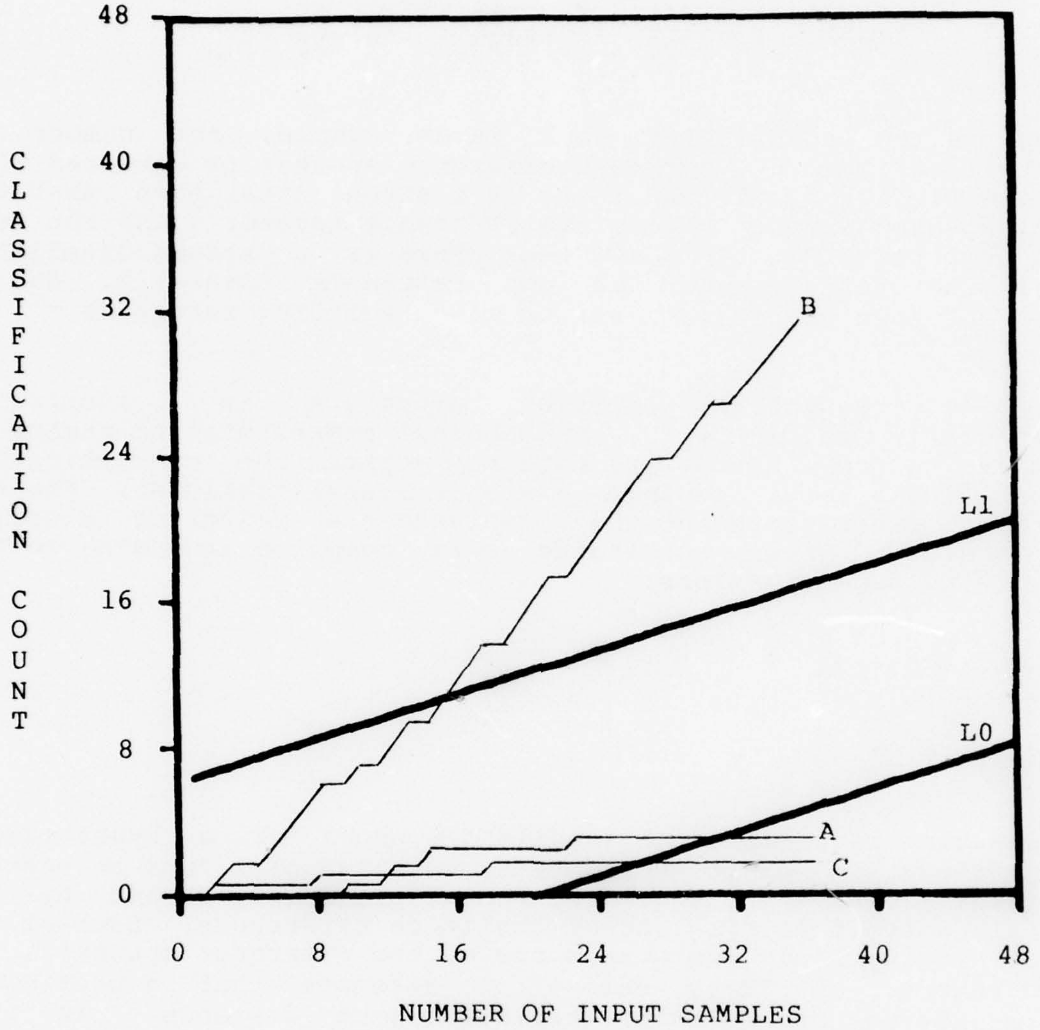


Figure 17. Graphical illustration of the sequential analysis decision procedure for an unknown speaker. The two parallel lines represent acceptance and rejection decision thresholds. The stepped lines represent the classification counts for different reference speakers. References A and C are rejected and reference B is accepted as being the unknown speaker.

4.5 Experiments Using Sequential Analysis

The intermediate results of the text-independent speaker identification experiment without vowel recognition (which resulted in the confusion matrix in Figure 16) were plotted for each test speaker, thereby giving 20 sequential analysis plots. These are illustrated in Figures 18-37, one for each test speaker. The test samples from the unknown speakers were presented in discourse order so the plots represent a practical situation of sequential testing as vowels are detected, but not recognized, in a speech recording. The acceptance and rejection decision lines have been plotted in each figure so that the results of a sequential decision procedure can be observed. The specifications of the decision lines are:

P[FR]=.05
P[FA]=.05
P1=.35
P0=.25.

A detailed description of the first sequential plot (Figure 18) will be given, followed by a discussion of the overall sequential analysis results for all 20 test speakers.

Figure 18 illustrates the sequential intermediate results of sample classifications when the unknown speaker is JOE and there are 20 possible speakers. Each reference speaker is represented by a line which depicts the number of times a particular reference had the minimum distance (out of 20) to an input test sample. Under ideal conditions, when speaker JOE is tested, all test samples would be classified as reference speaker JOE, and there would be one 45 degree line for JOE, with the 19 other references lying along the baseline. Unfortunately, this rarely occurs in real-world situations. Furthermore, it can be seen that the JOE reference always had the highest classification count. Reference speaker LLP consistently had the second highest classification count and reference JC was consistently ranked third. The classification counts for the other reference speakers were sufficiently low that they have minimal effect.

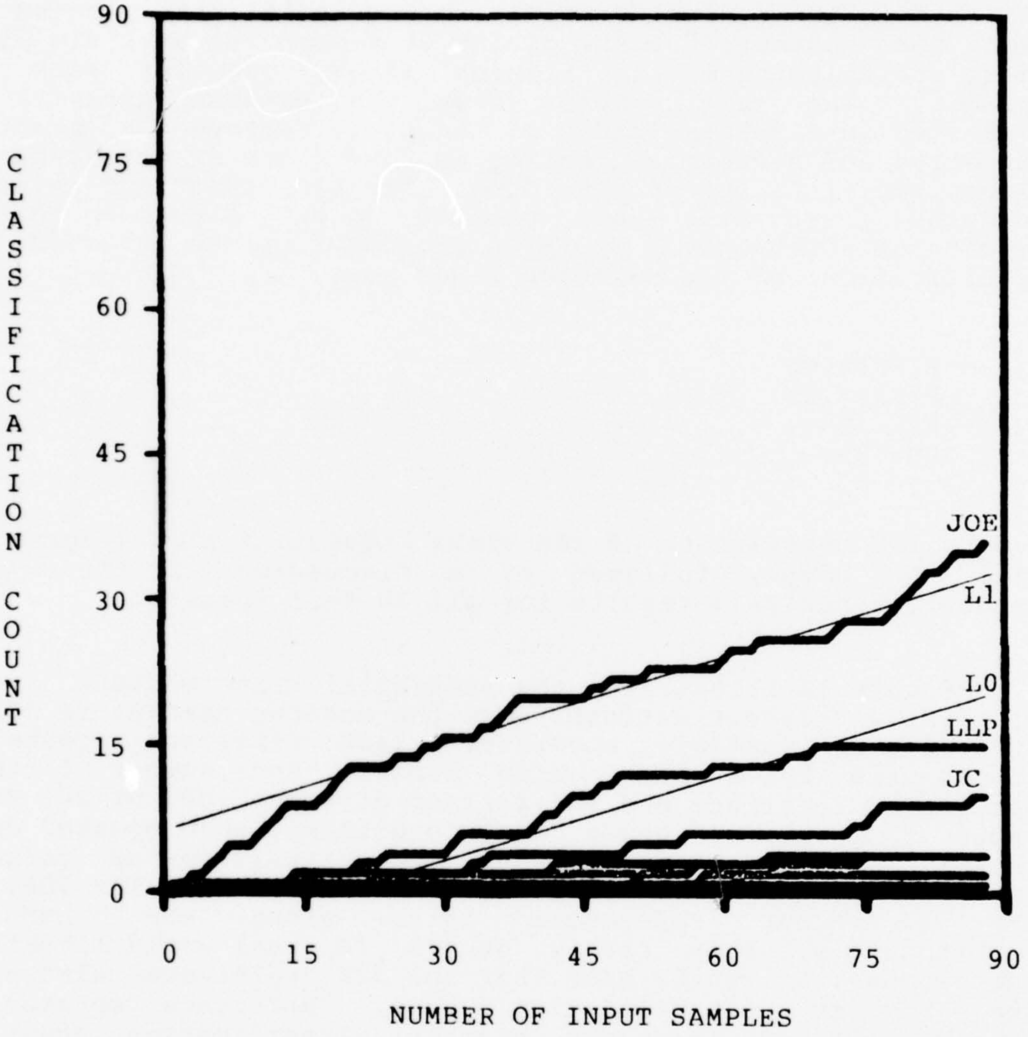


Figure 18. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JOE. Reference samples from data set 5, test samples from data set 6.

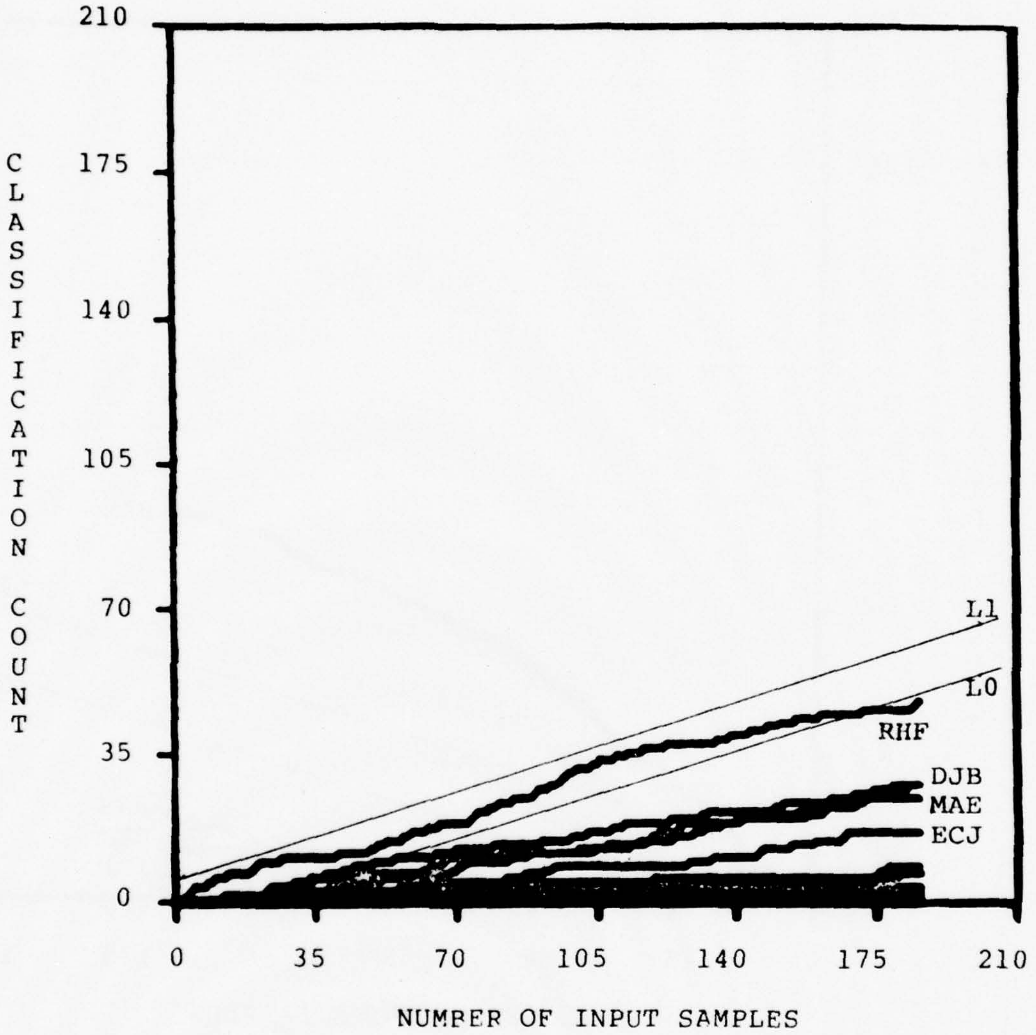


Figure 19. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker RHF. Reference samples from data set 5, test samples from data set 6.

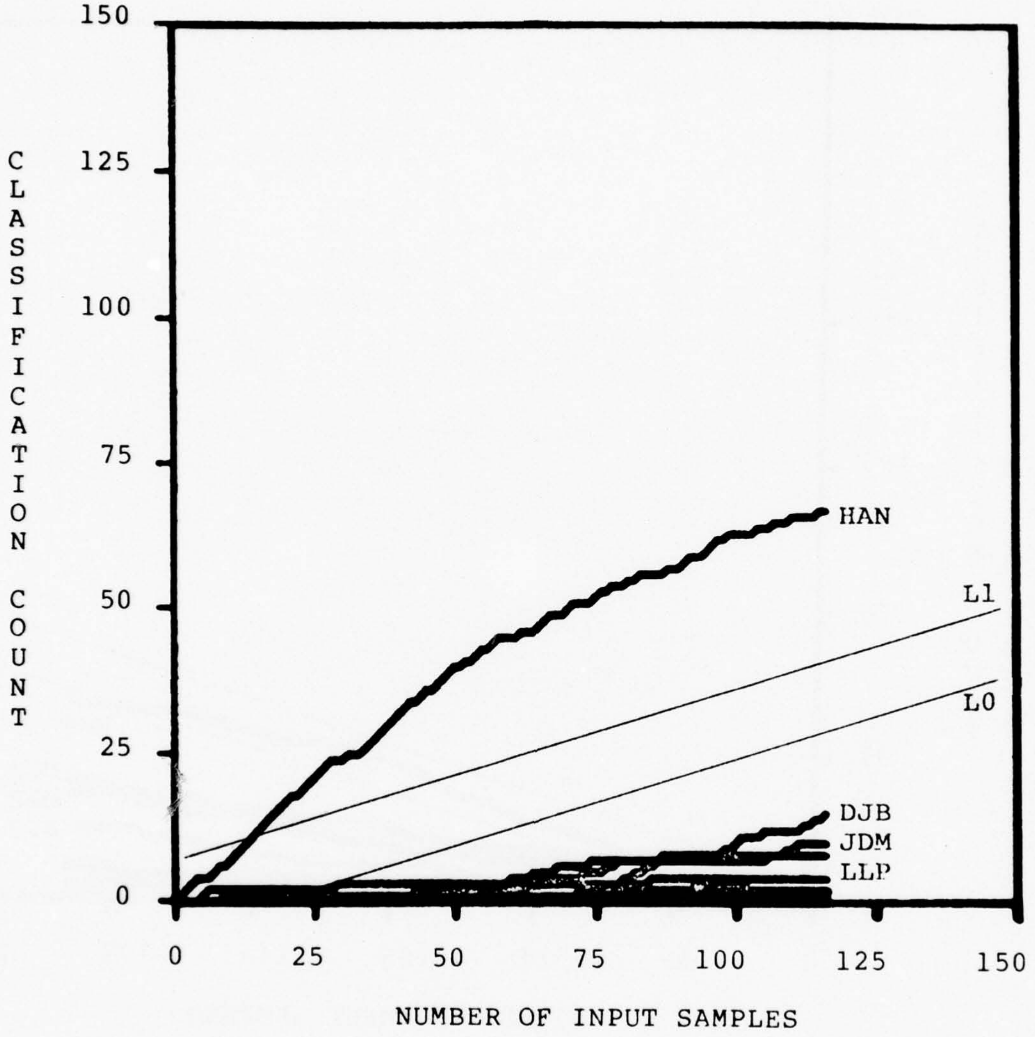


Figure 20. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker HAN. Reference samples from data set 5, test samples from data set 6.

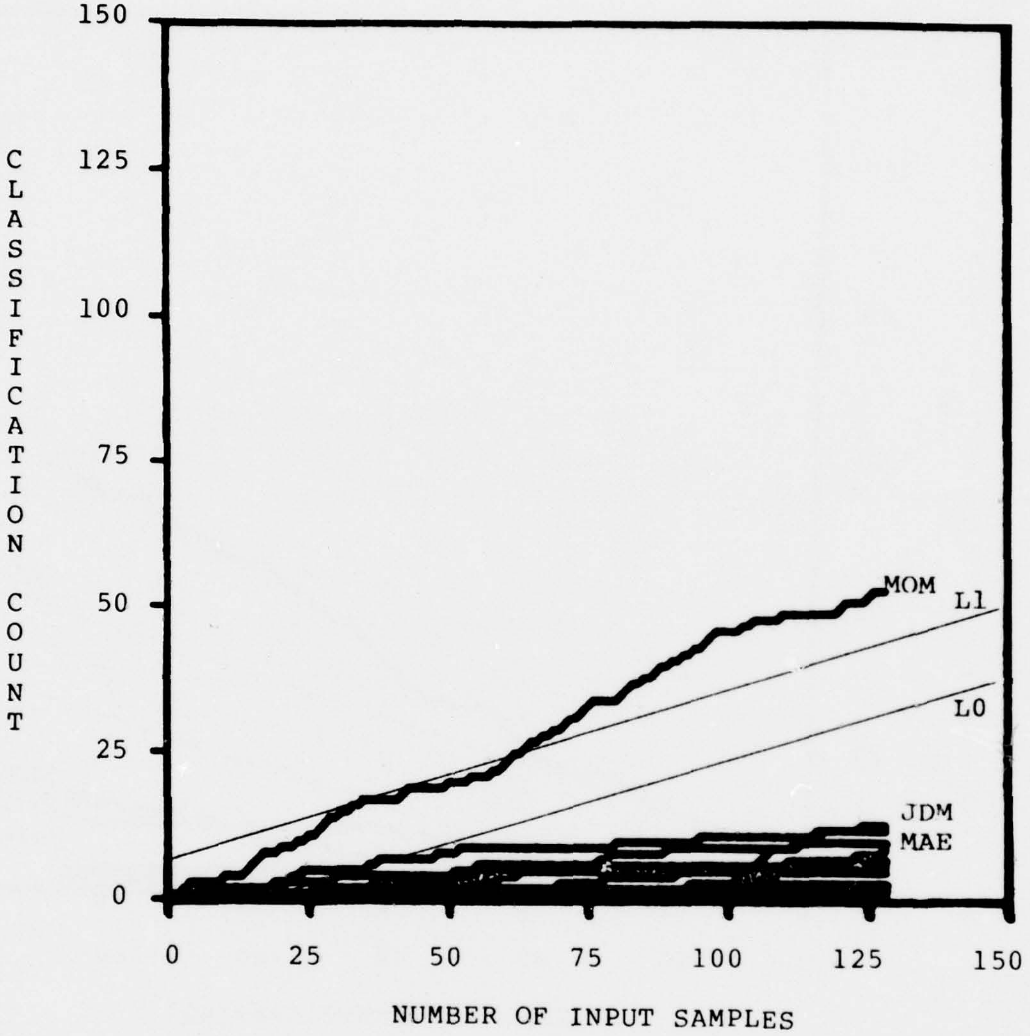


Figure 21. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker MOM. Reference samples from data set 5, test samples from data set 6.

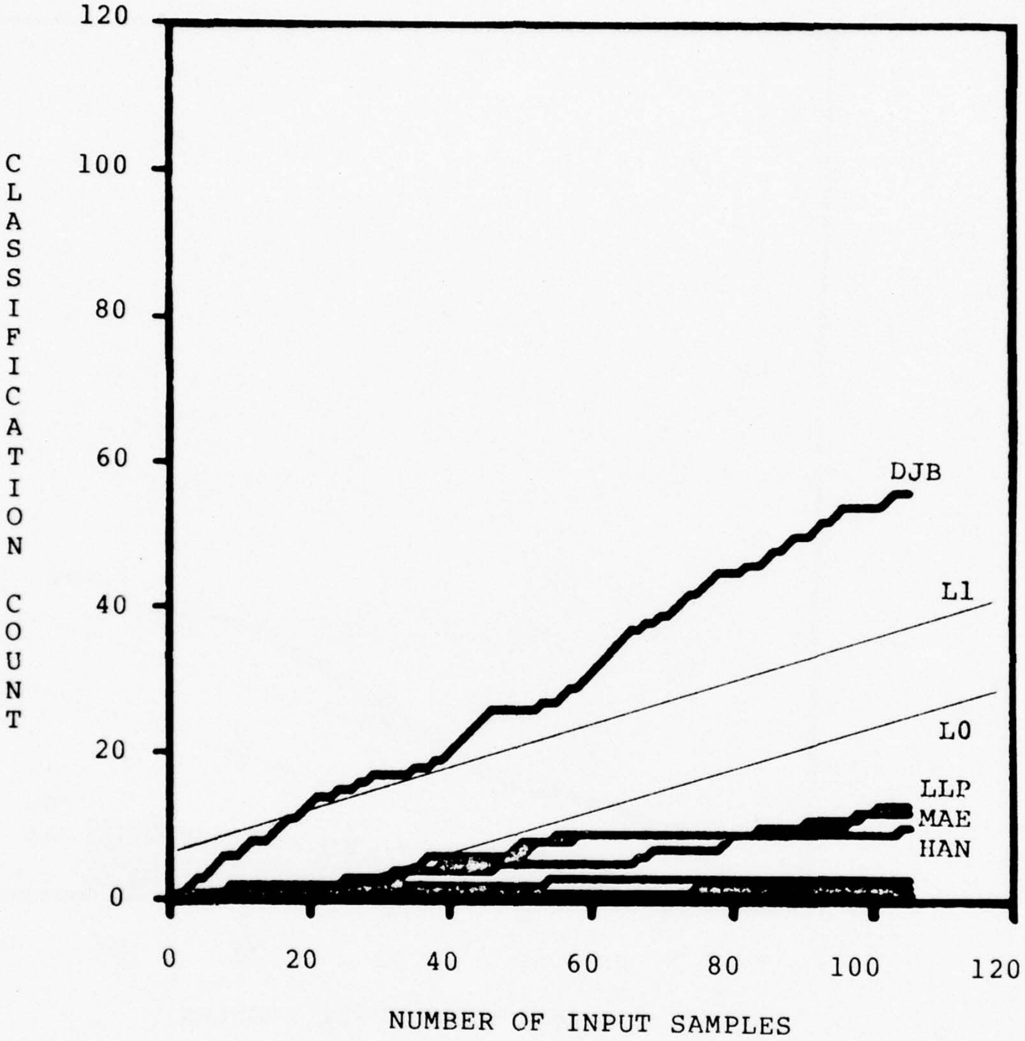


Figure 22. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker DJB. Reference samples from data set 5, test samples from data set 6.

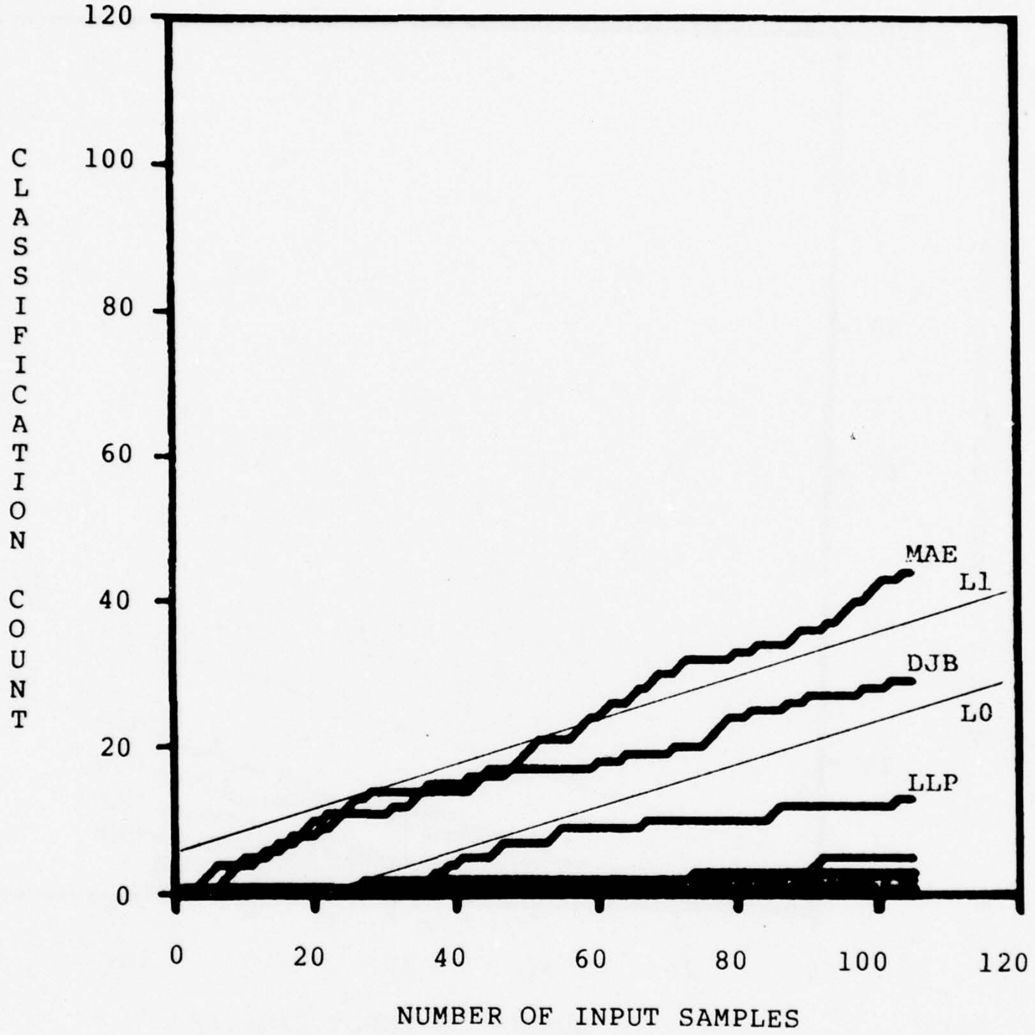


Figure 23. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker MAE. Reference samples from data set 5, test samples from data set 6.

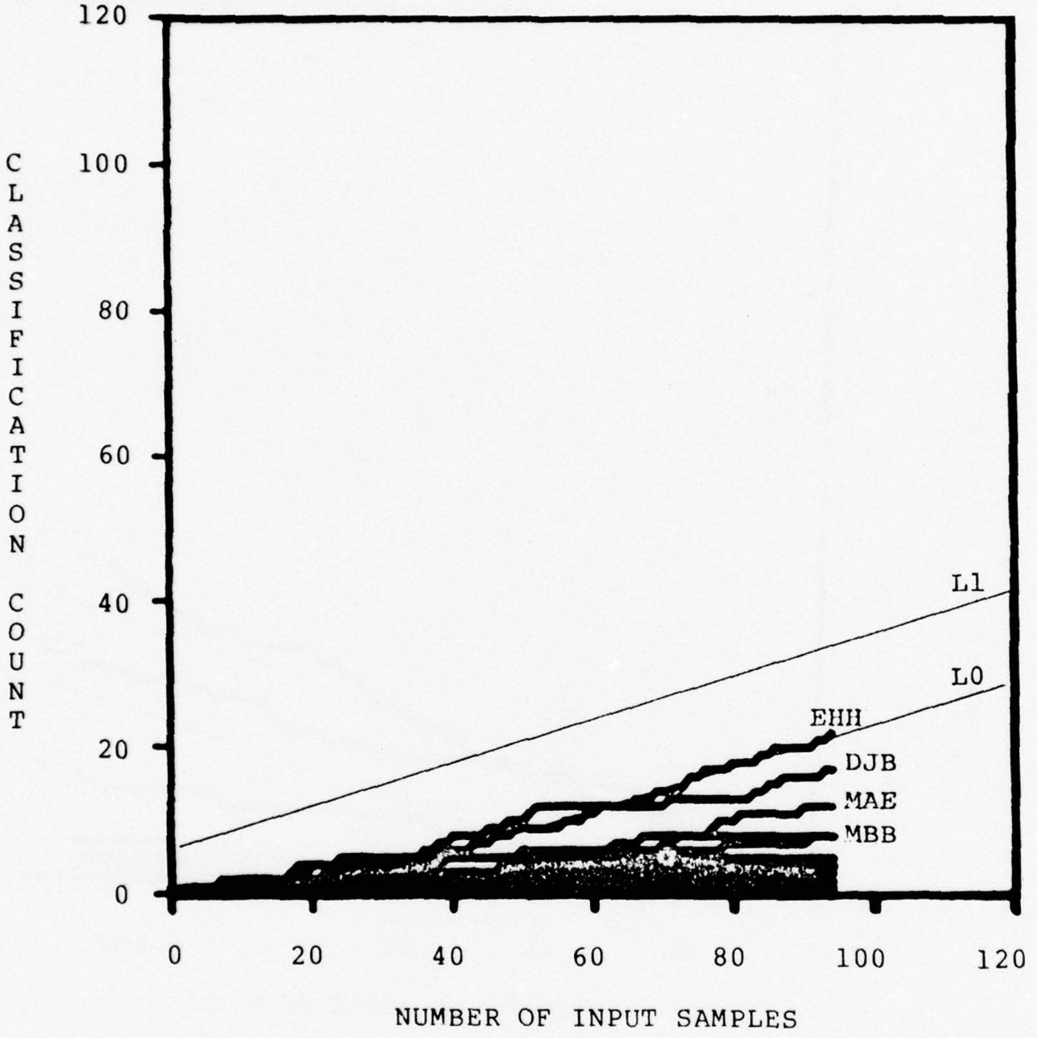


Figure 24. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker EHH. Reference samples from data set 5, test samples from data set 6.

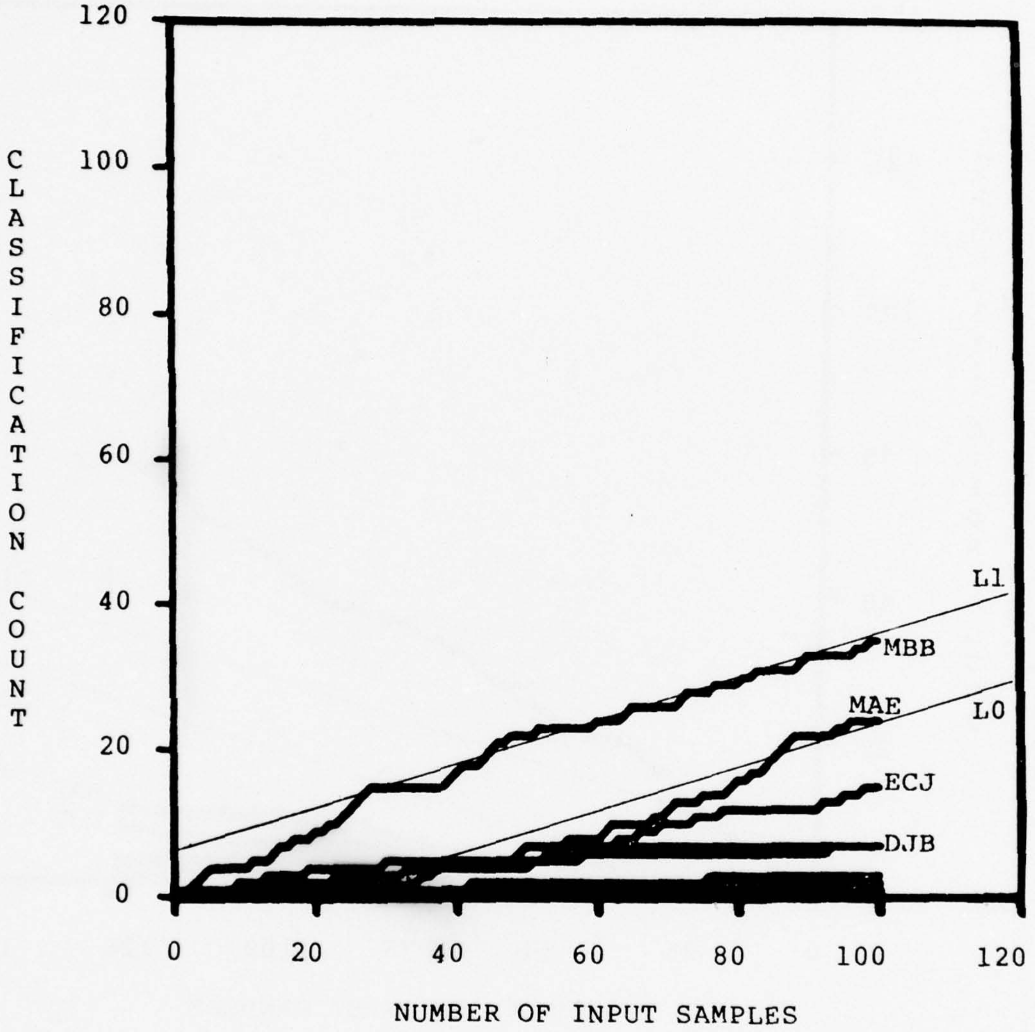


Figure 25. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker MBB. Reference samples from data set 5, test samples from data set 6.

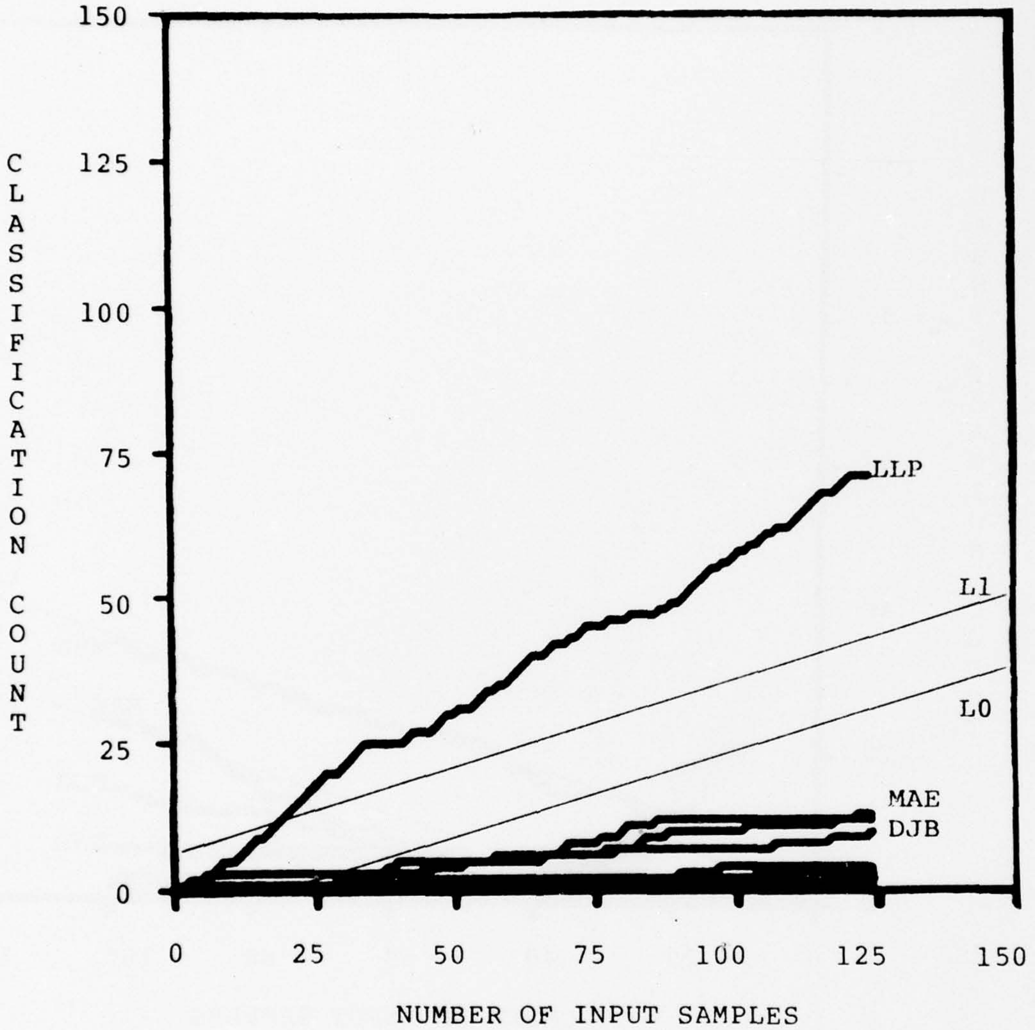


Figure 26. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker LLP. Reference samples from data set 5, test samples from data set 6.

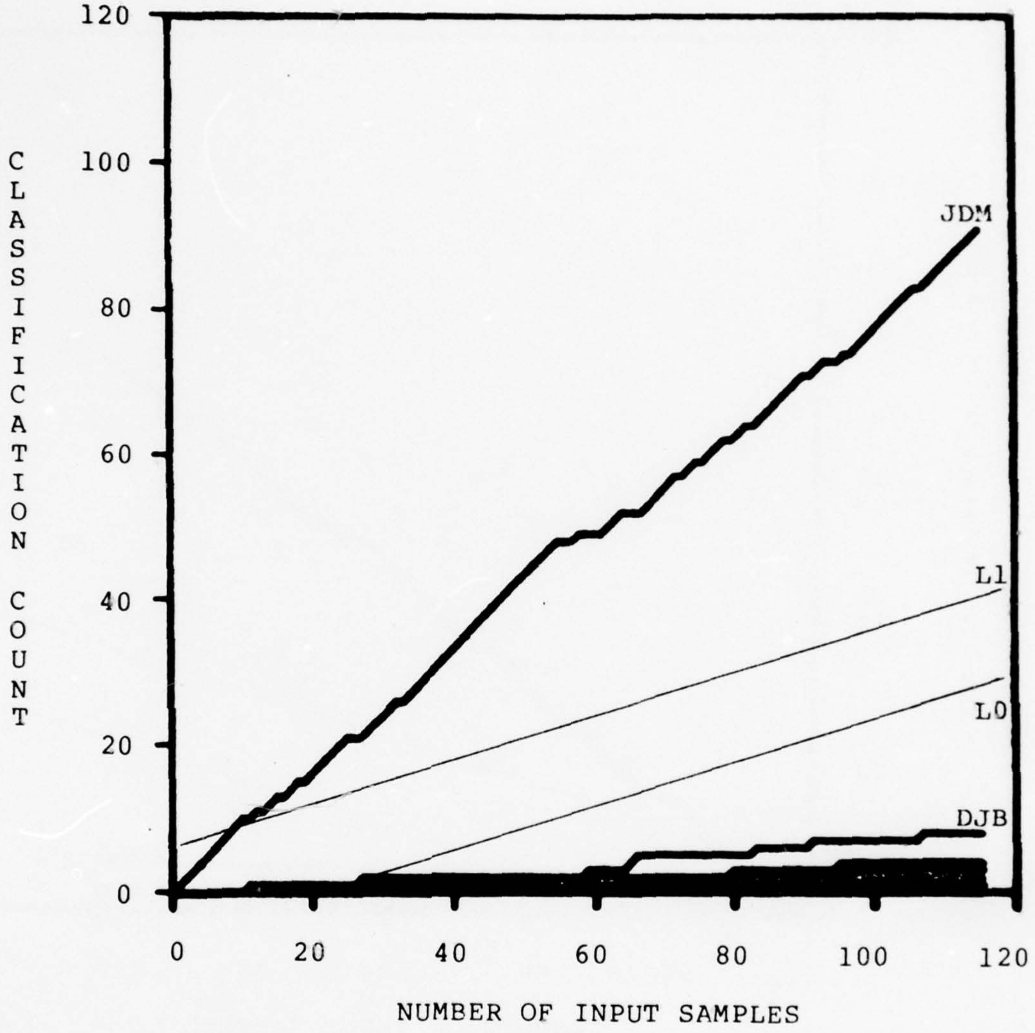


Figure 27. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JDM. Reference samples from data set 5, test samples from data set 6.

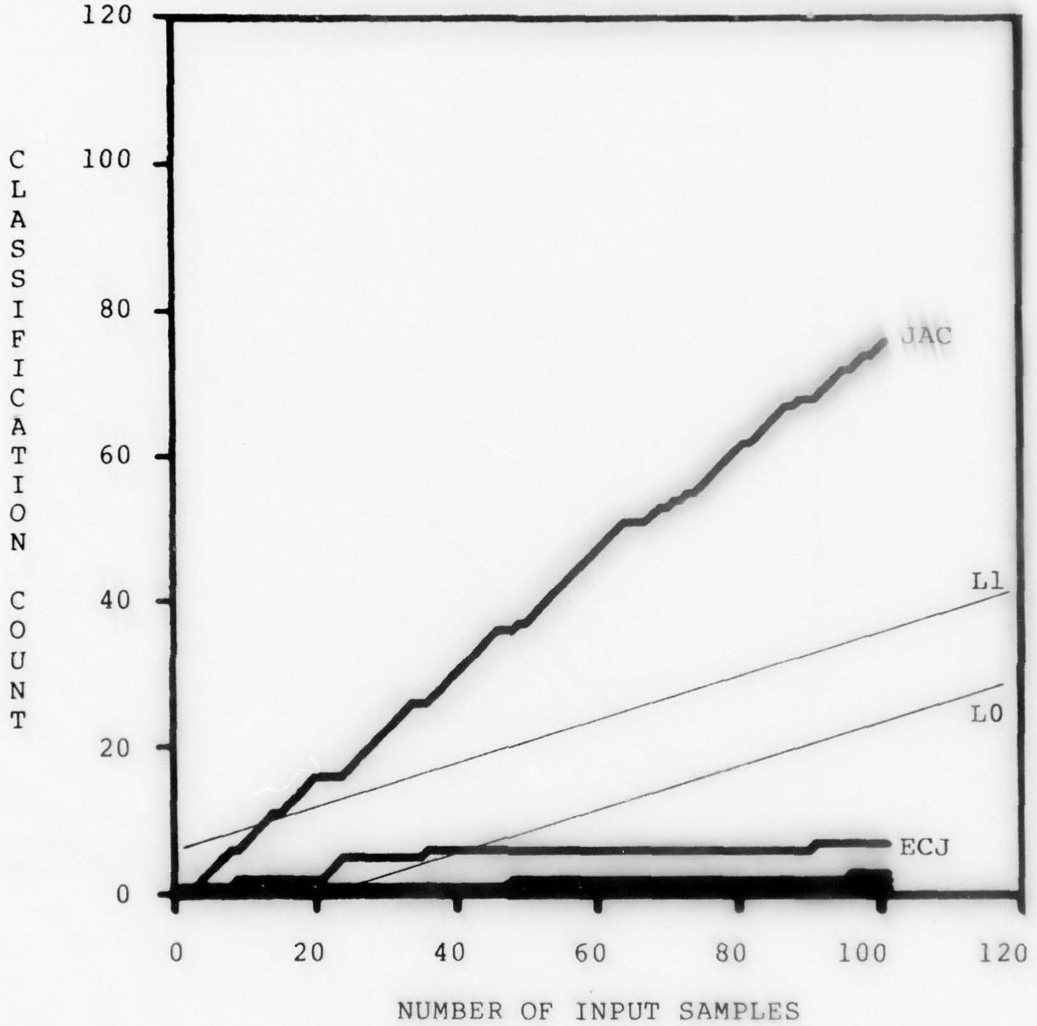


Figure 28. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JAC. Reference samples from data set 5, test samples from data set 6.

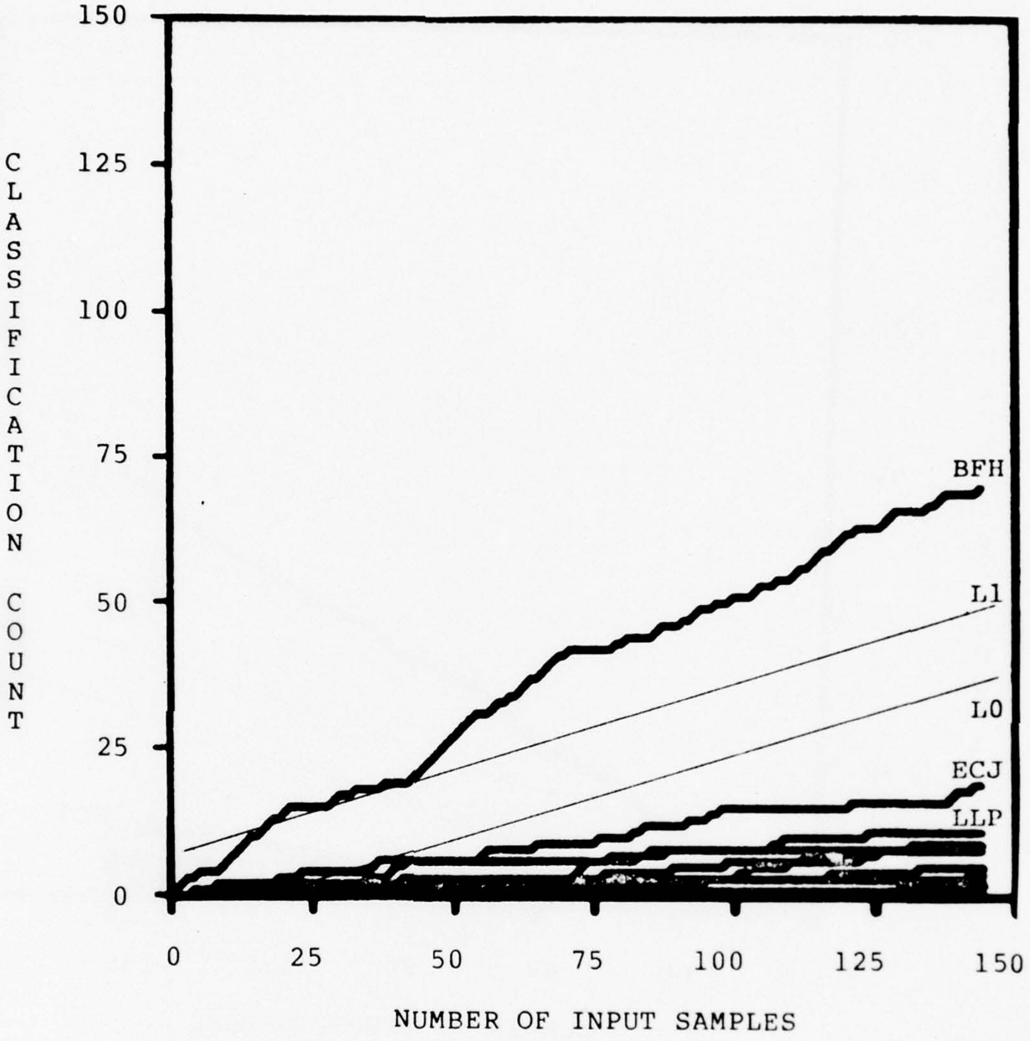


Figure 29. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker BFH. Reference samples from data set 5, test samples from data set 6.

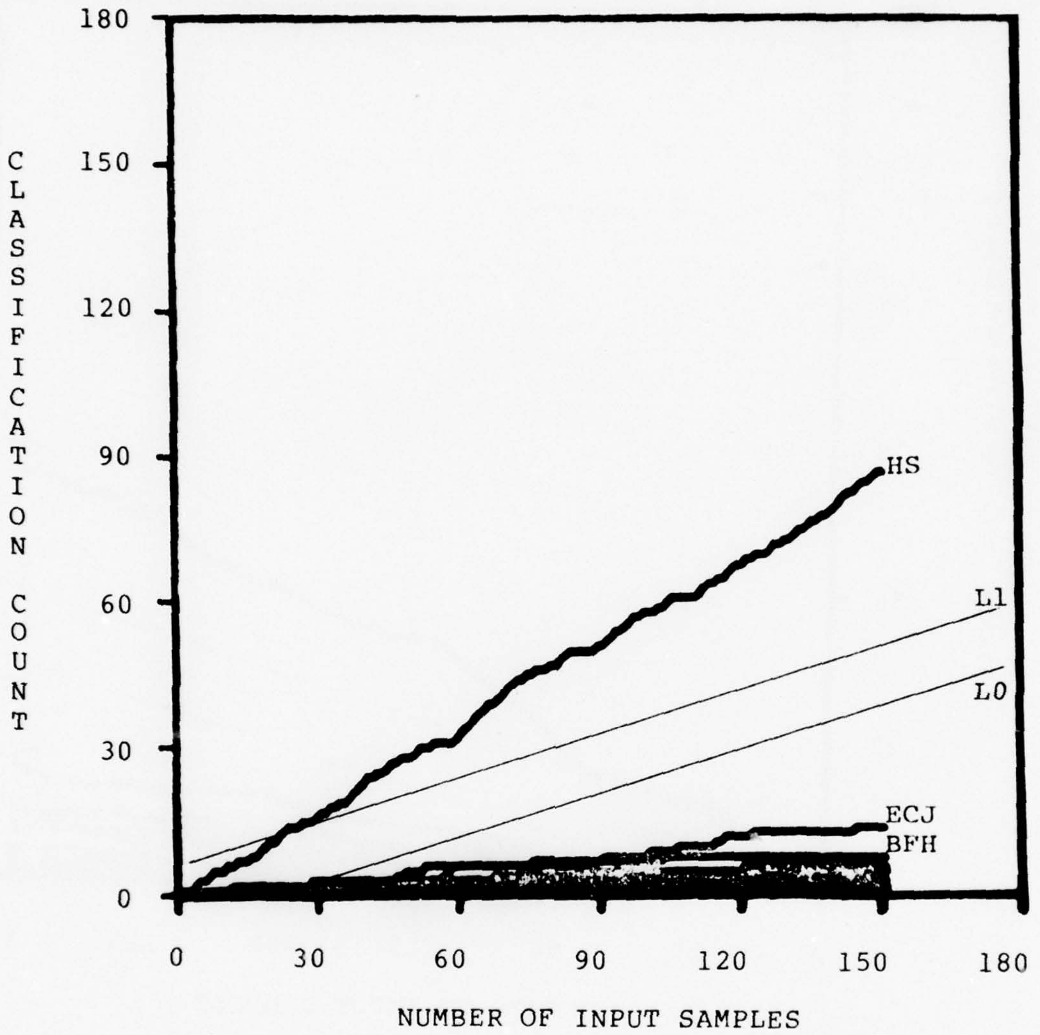


Figure 30. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker HS. Reference samples from data set 5, test samples from data set 6.

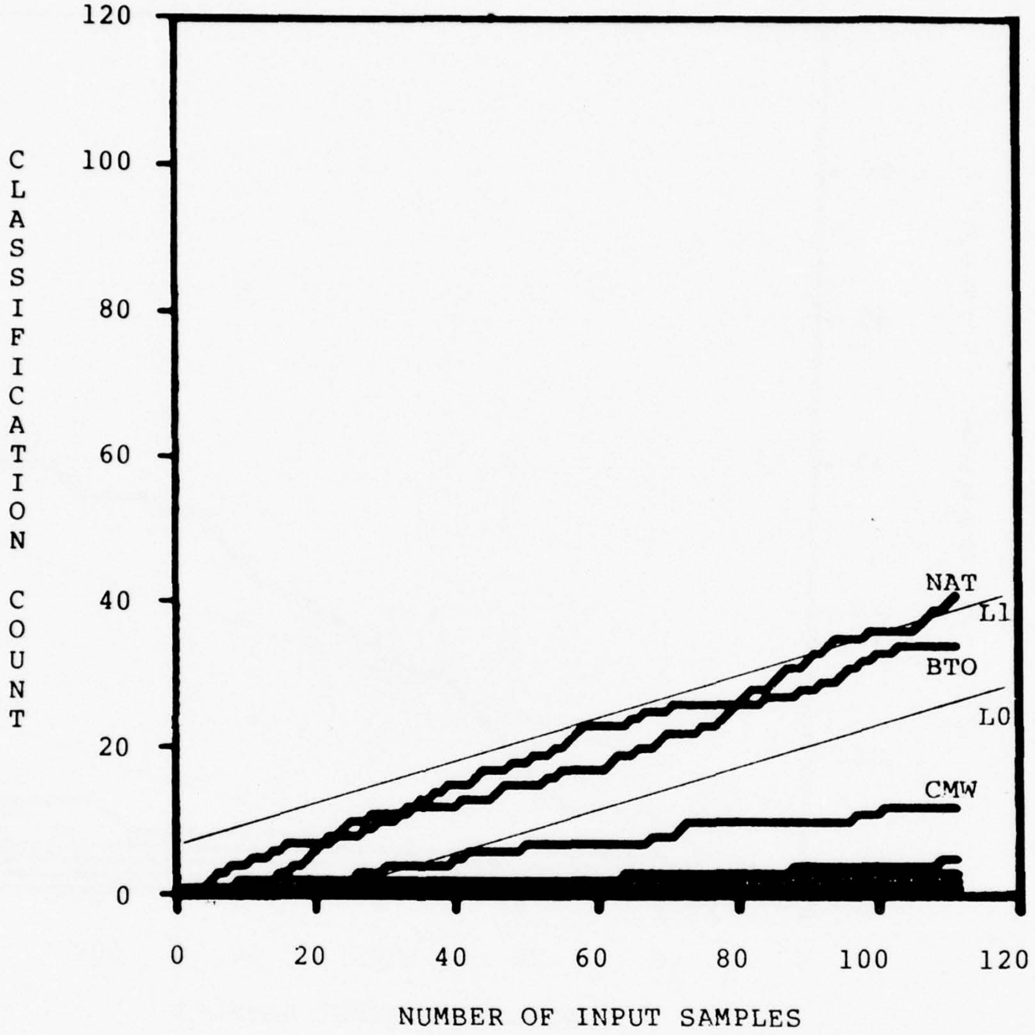


Figure 31. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker BTO. Reference samples from data set 5, test samples from data set 6.

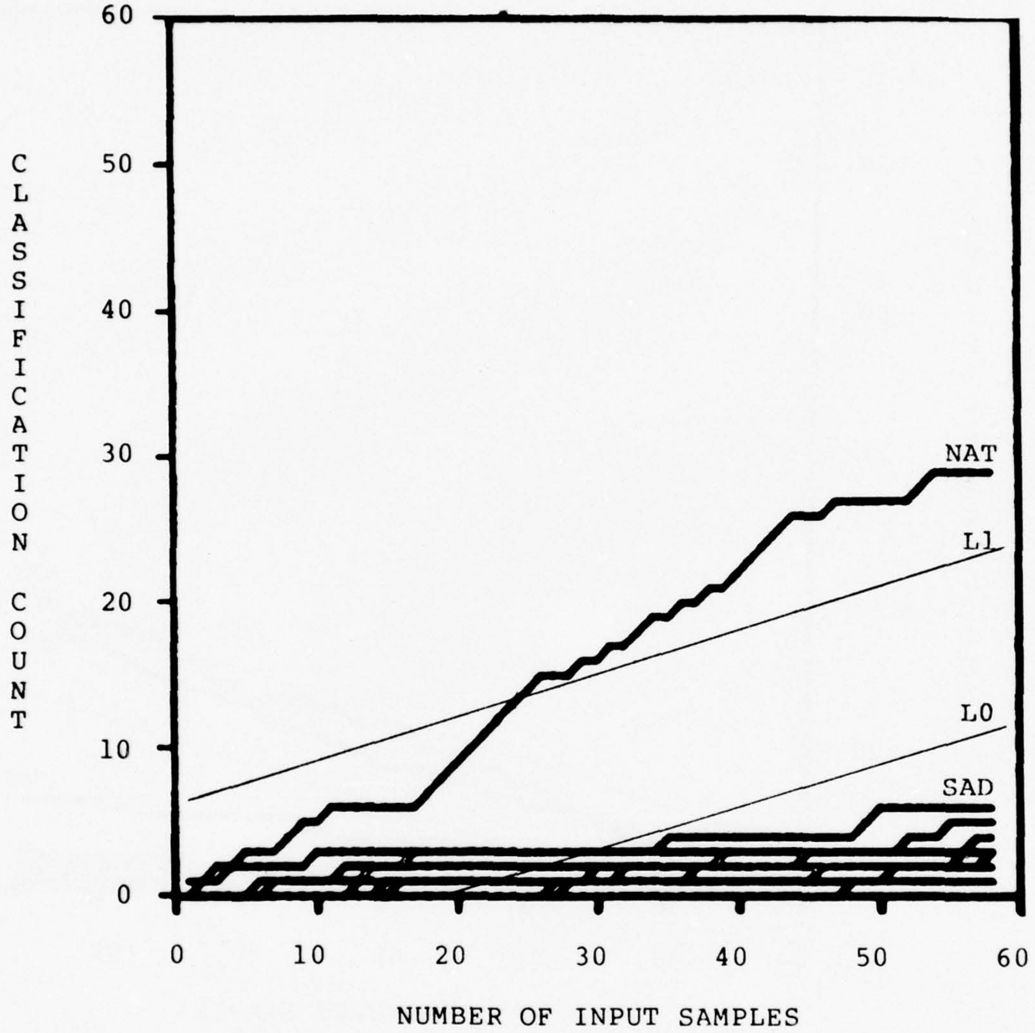


Figure 32. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker NAT. Reference samples from data set 5, test samples from data set 6.

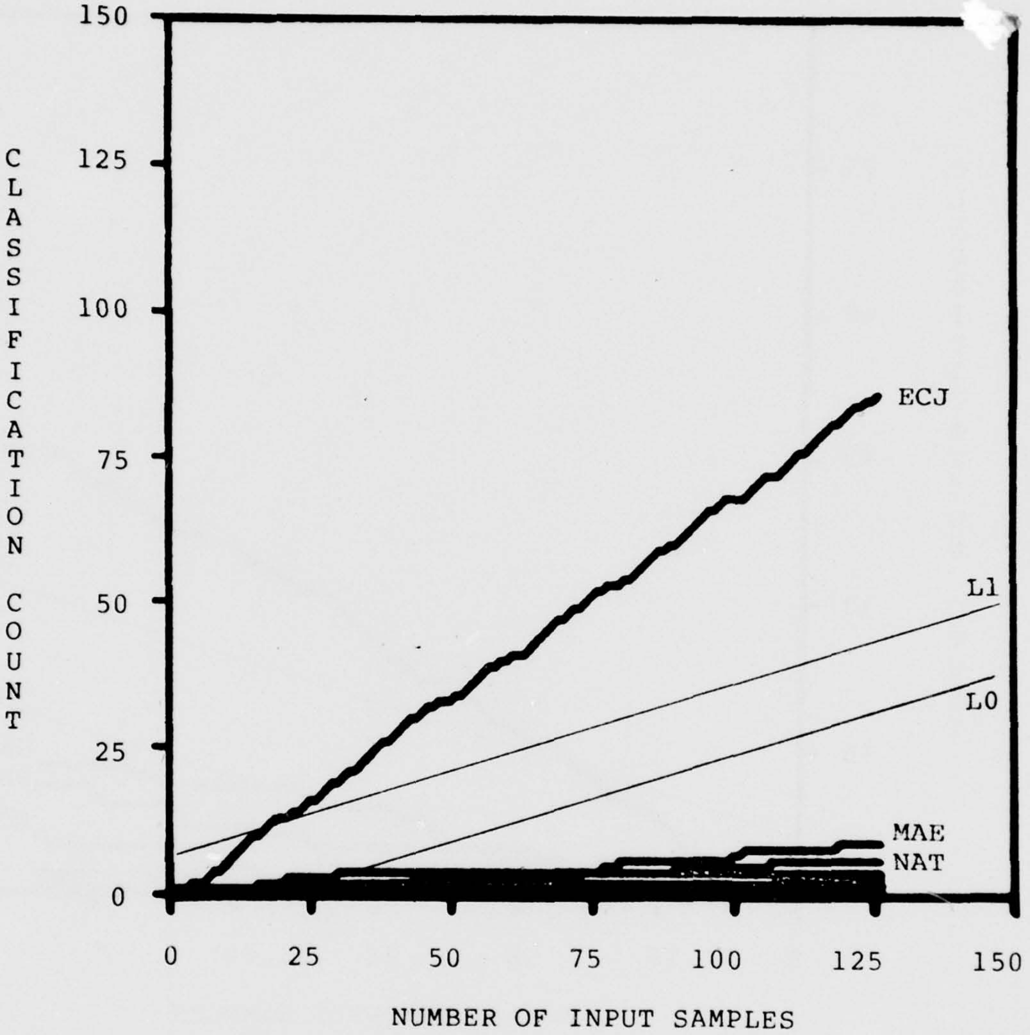


Figure 33. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker ECJ. Reference samples from data set 5, test samples from data set 6.

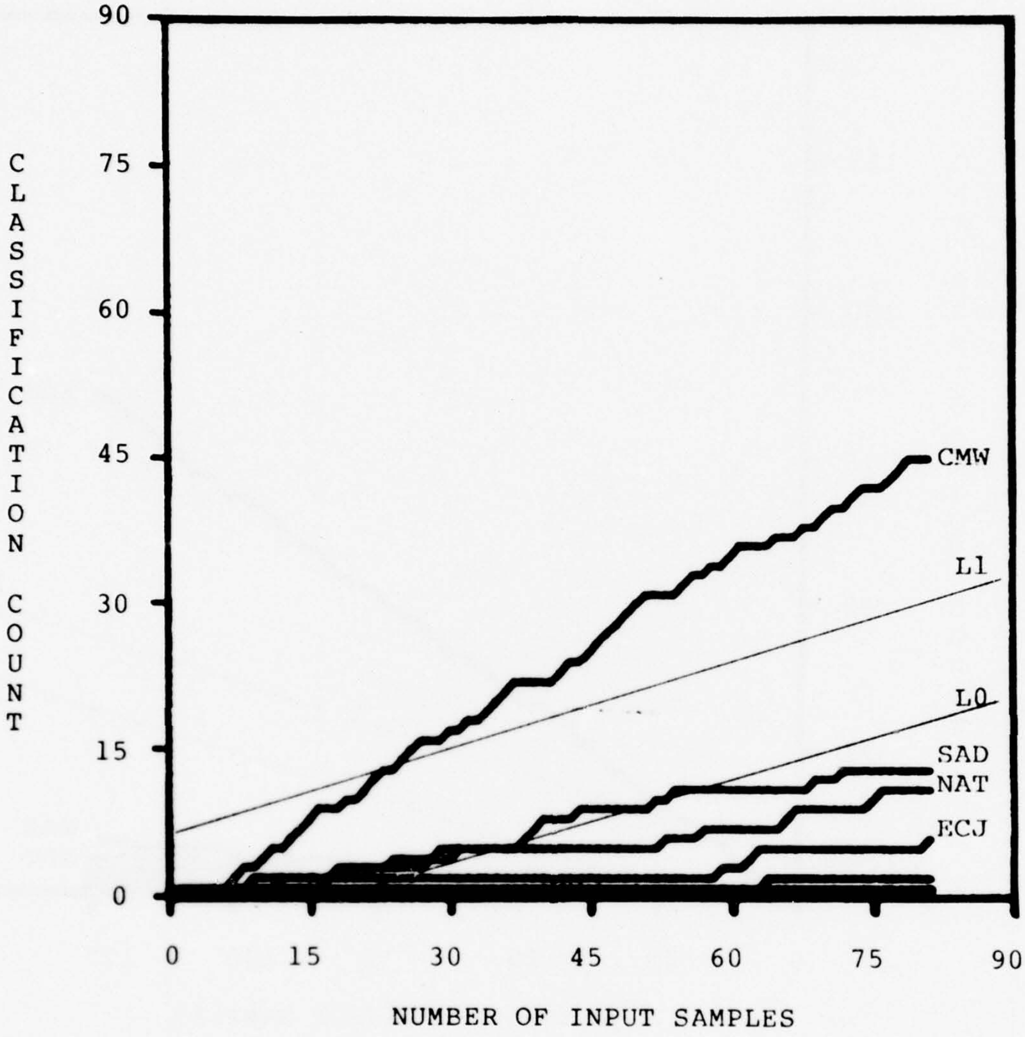


Figure 34. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker CMW. Reference samples from data set 5, test samples from data set 6.

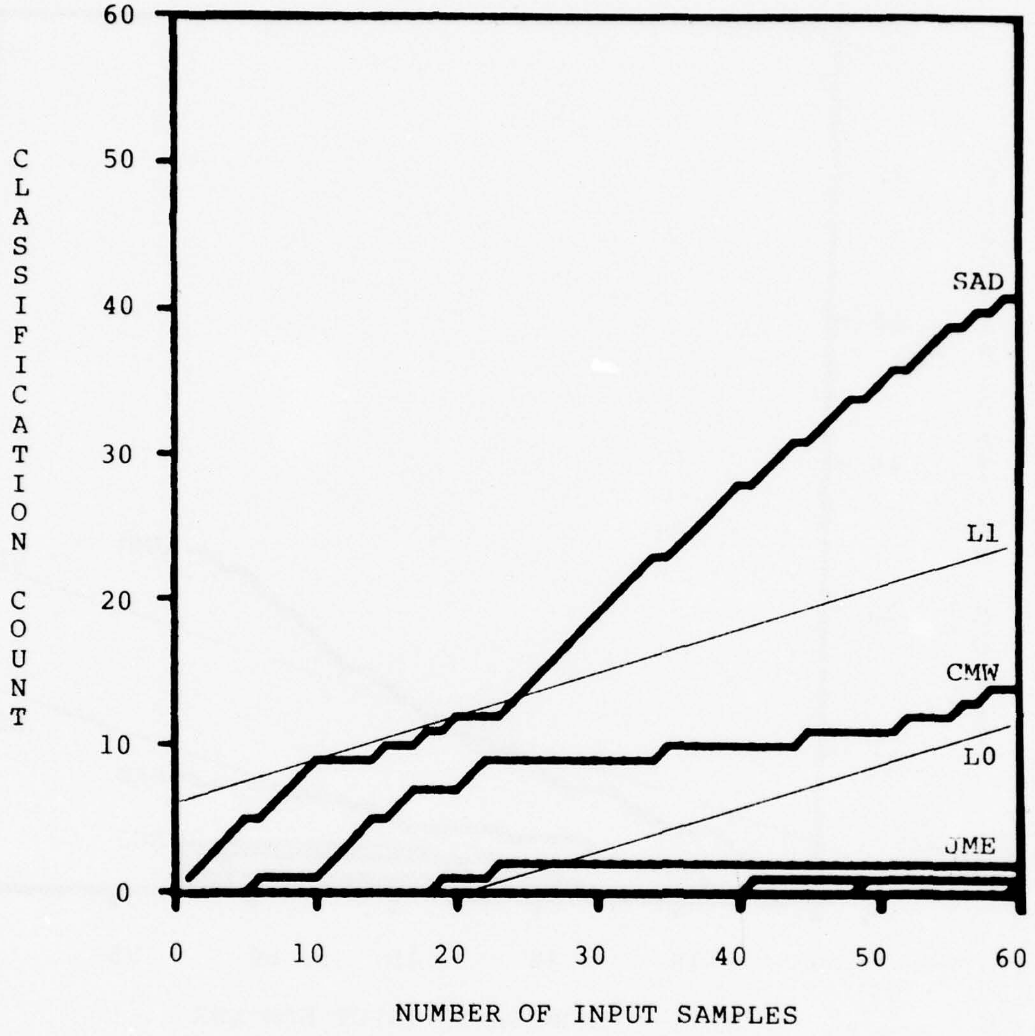


Figure 35. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker SAD. Reference samples from data set 5, test samples from data set 6.

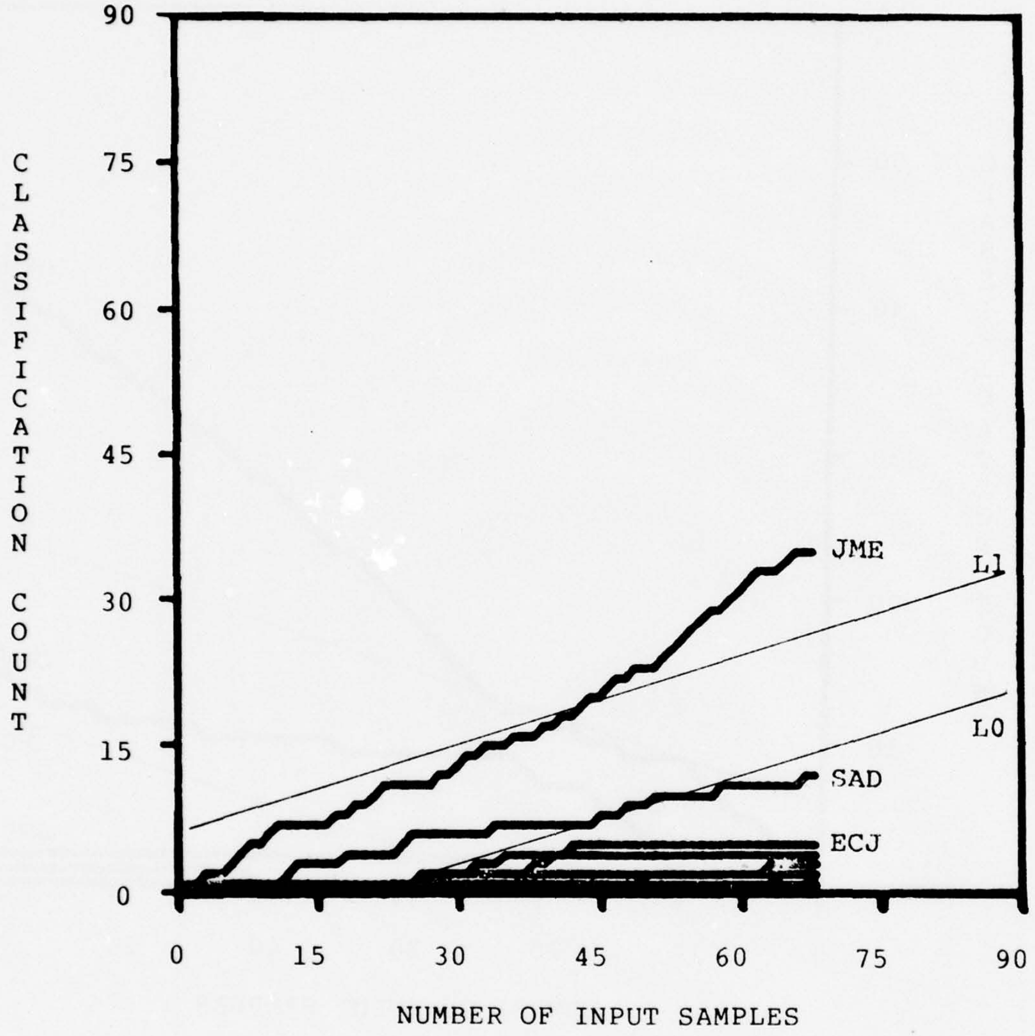


Figure 36. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JME. Reference samples from data set 5, test samples from data set 6.

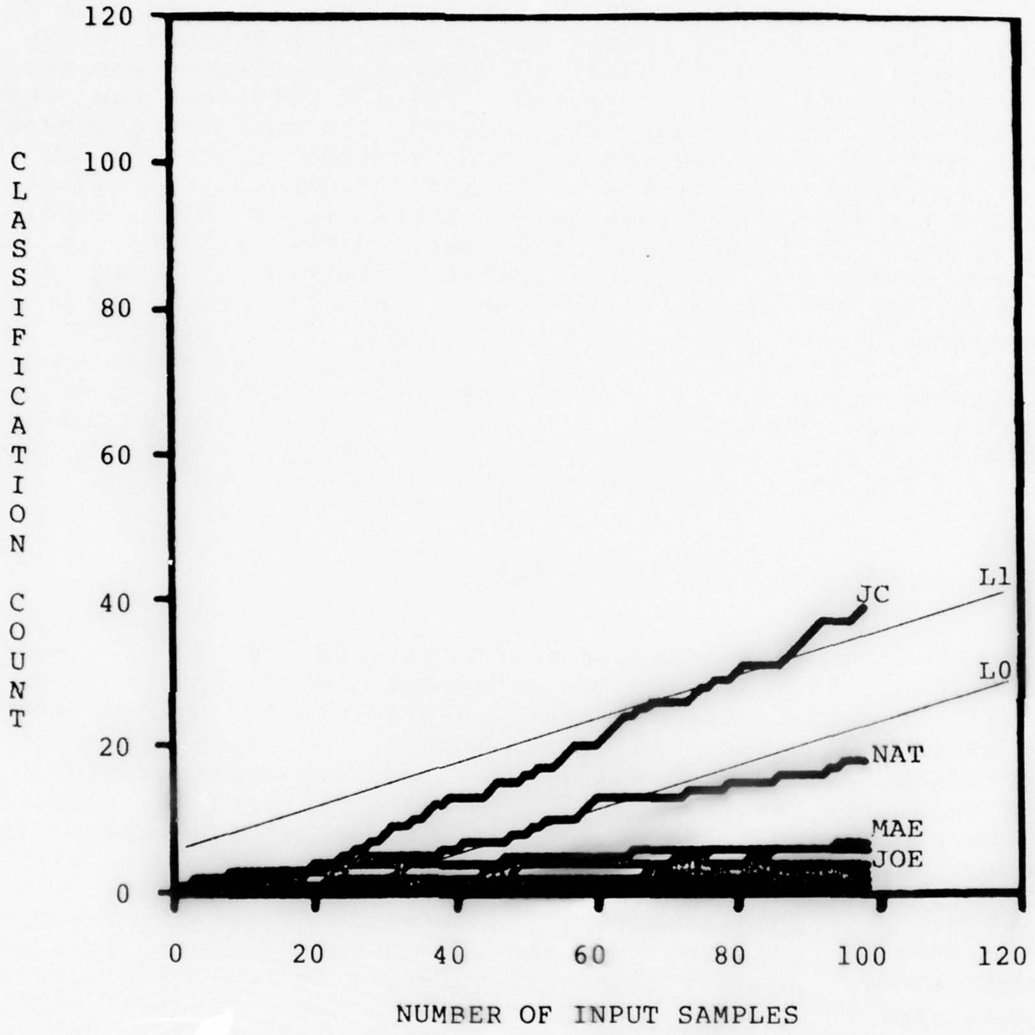


Figure 37. Plot of sequential analysis results for text-independent speaker identification without vowel recognition. Test speaker JC. Reference samples from data set 5, test samples from data set 6.

Relating to Figure 18, the classification count for all the references except JOE and LLP have crossed the rejection line within thirty samples or less. According to the defined decision criteria, there must be at least 20 test samples before a reference can be rejected. For the 18 references which could be rejected there is a .05 probability that the rejection is in error. The sequential result for speaker LLP follows close to the rejection threshold until 40 vowel samples have been tested, at which time it can be rejected. The JOE reference was chosen sufficiently often that it crossed the acceptance threshold after only 20 vowel samples had been tested. Since JOE was the only reference to cross the acceptance threshold, and all others crossed the rejection threshold, there is a high level of confidence in identifying the test speaker as JOE. In fact, according to the decision criteria, there is only a .05 probability of false acceptance. In this case, a correct identification decision could be made after only 19 vowel samples, or 25 seconds. While the detailed sequential analysis results for each speaker tested are shown in Figures 18-37, the results are summarized in Table 9. This table lists the decision results, the number of vowel samples required for a decision and the corresponding time to reach a decision.

4.5.1 Sequential analysis results

The sequential analysis procedure resulted in 17 correct identifications out of 20 attempts. When speaker JDM was tested, it only required 9 vowel samples, or 5.5 seconds to make a decision. The longest amount of time to make an identification decision was 74.2 seconds, 89 vowel samples, and this was for test speaker JC. For the 17 correctly identified speakers, the average time for decision was 29.4 seconds. It is clear that the time to decision is speaker dependent, with the more distinct speakers requiring fewer test samples. It is interesting to note, however, that because of speaker-to-speaker differences in speaking rate and the amount of pause, equivalent vowel counts do not necessarily result in equivalent decision times. For example, it required only 25 vowel samples each to reach a correct identification decision when speaker NAT, CMW, and SAD were tested. However, the decision times for NAT and CMW were 34.6 seconds and 36.8 seconds respectively, whereas speaker SAD spoke much slower and therefore her 25 vowel samples spanned 52.9 seconds. These results illustrate the utility of the sequential analysis procedures in making a decision in as short a time as possible, and its ability to provide a level of confidence in the decision through the specification of operational error criteria and probabilistic expectation.

Table 9. Results of text-independent speaker identification without vowel recognition, using a sequential analysis decision procedure.

TEST SPEAKER	IDENTIFICATION DECISION	NUMBER SAMPLES TO DECISION	TIME TO DECISION
JOE	JOE	19	25.0 sec
RHF	All Rejected	185	*
HAN	HAN	14	12.0 sec
MOM	MOM	62	35.3 sec
DJB	DJB	19	23.1 sec
MAE	MAE	59	43.4 sec
EHH	All Rejected	66	*
MBB	MBB	45	26.3 sec
LLP	LLP	19	12.3 sec
JDM	JDM	9	5.5 sec
JAC	JAC	14	15.4 sec
BFH	BFH	16	10.5 sec
HS	HS	27	14.8 sec
BTO	NAT**	93	72.3 sec
NAT	NAT	25	34.6 sec
ECJ	ECJ	18	10.9 sec
CMW	CMW	25	36.8 sec
SAD	SAD	25	52.9 sec
JME	JME	43	66.8 sec
JC	JC	89	74.2 sec

* Special Case (see section 4.5.2)
** False Acceptance

Using the sequential analysis decision procedure, there was one case of false acceptance, and this occurred for test speaker BTO, who was identified as NAT. This particular error was anticipated because test speaker BTO was the only one who did not get the highest classification count in the confusion matrix results (refer back to Figure 16). However, an examination of the sequential data for test speaker BTO (See Figure 31) shows that the identification decision may not be complete. Even though reference NAT has crossed the acceptance threshold, reference BTO is close by and still in the no-decision region. Thus it could be that the sequential decision procedure should allow for selecting multiple potential candidates. This would certainly have to be the case if more than one reference crossed the acceptance threshold.

4.5.2 Special cases

There were two cases in the sequential analysis decision experiment where all references were rejected. This situation occurred when speakers RHF and EHH were tested. The classification counts were sufficiently distributed that eventually even the true reference speaker was rejected. Actually, since the experiment was set up as closed-choice tests, rejection of all candidates is not an allowable decision. A forced choice decision could still be made based upon the highest classification count. If such a decision was enforced, then for test speaker RHF, it can be seen from the sequential results in Figure 31 that reference RHF would qualify as the highest potential candidate, although such a decision would not have the strength as those which met the acceptance criterion of the acceptance threshold. A higher level of confidence could be added to that decision, however, by comparing the minimum distance values for RHF and the minimum distance values for the reference speaker with the second highest classification count, namely DJB. Based upon 189 test samples, a t-Test on the means of the minimum distances for references RHF and DJB shows that they are significantly different at the .01 level. Thus, a single choice of RHF would be justified, and would be correct as well. In the case of test speaker EHH, the sequential results (Figure 24) illustrate that the reference EHH has the highest classification count after 94 test samples and reference DJB has the second highest classification count. A t-Test on the minimum distances for references EHH and DJB shows that their means are not significantly different at the .01 level. A forced decision would therefore have to conclude that the test speaker might be either EHH or DJB, and thus the true speaker is included in the selected pair.

4.6 Test Data from Two-Weeks Later

Two text-independent speaker identification tests were performed using the test samples from speakers JDM and BFH recorded after a two-week interval. The samples were tested in discourse order, and the overall results of the sample classifications are shown in the confusion matrix of Figure 38. For test speaker JDM, 98 of the 133 test samples were classified as JDM. It is clear from these results and those of previous experiments, that JDM is a very robust speaker who seems to be distinct from the others. For test speaker BFH, 38 of the 115 test samples were classified as BFH. This classification count is not as significant as it was for speaker JDM.

The significance of the results from the two-week test data can be evaluated by means of the sequential analysis procedure. Table 10 gives a summary of the decision results. A graphical illustration of the sequential results for speaker JDM are shown in Figure 39 and for speaker BFH in Figure 40. The acceptance and rejection thresholds have the following specifications:

P[FR]=.05
P[FA]=.05
P1=.35
P0=.25.

When speaker JDM is tested, the JDM reference crosses the acceptance threshold after 28 vowel samples, or 18 seconds of test data. All other references are rejected. Thus after a time span of two weeks, speaker JDM can still be corretly identified with a high level of confidence.

The sequential results for the BFH test samples show that 19 of the 20 references are rejected, but the BFH reference remains in the no-decision region until the test samples are exhausted. In a forced-choice test, a correct identification decision would still be made because the BFH reference has the highest classification count and the difference between it and the next nearest reference is significant at the .01 level (difference between means of minimum distance). The decision would be correct, but it would not have the level of confidence as that of the JDM test.

	JOE	RHF	HAN	MOM	DJB	MAE	EHH	MBB	LLP	JDM	JAC	BFH	HS	BTO	NAT	ECJ	CMW	SAD	JME	JC
JDM			6	3	15	6	2	1		98			1			1				
BFH	1		1		2	10		3	12			38	1	2	10	15	3	5		12

Figure 38. Confusion matrix for speaker identification experiment with two test speakers, where there was a two-week time interval between the reference samples and the test samples. Reference samples from data set 5 and test samples from data set 2.

Table 10. Results of text-independent speaker identification without vowel recognition, using a sequential analysis decision procedure.

TEST SPEAKER	IDENTIFICATION DECISION	NUMBER SAMPLES TO DECISION	TIME TO DECISION
JDM	JDM	28	18.0 sec
BFH	No Decision	115	90.0 sec

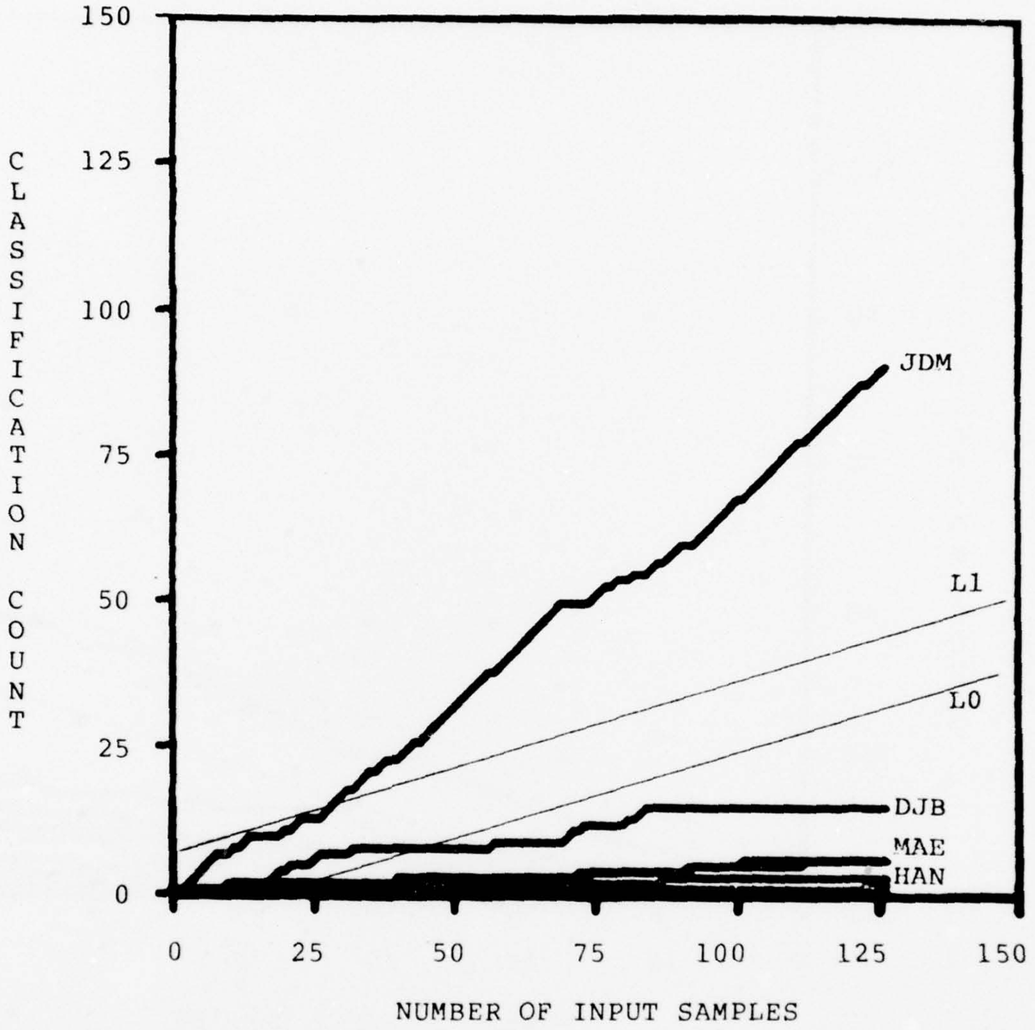


Figure 39. Plot of sequential analysis results for speaker identification without vowel recognition, and a two-week separation between test samples and reference data. Reference samples from data set 5 and JDM test samples from data set 2.

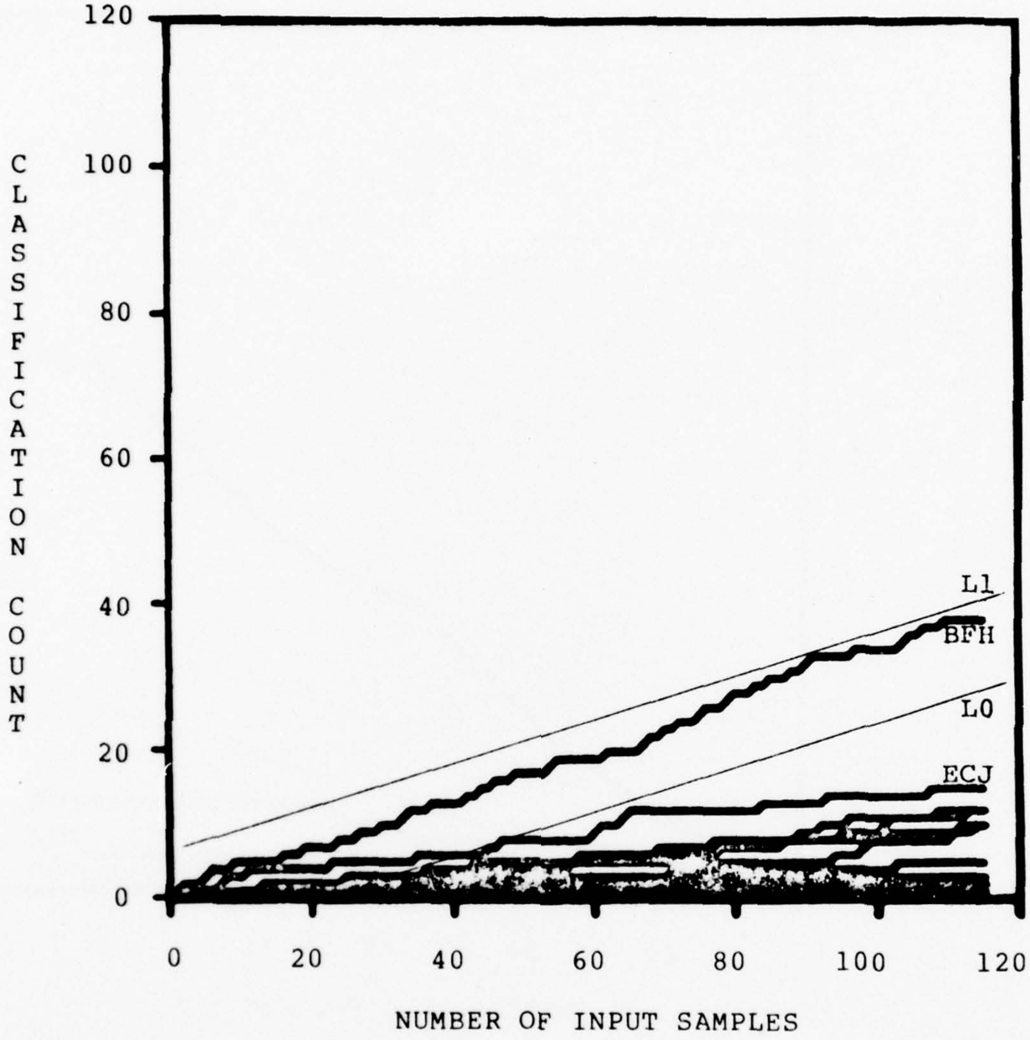


Figure 40. Plot of sequential analysis results for speaker identification without vowel recognition, and a two-week separation between test samples and reference data. Reference samples from data set 5 and BFH test samples from data set 2.

SECTION 5

SUMMARY AND RECOMMENDATIONS

5.1 Summary

A method for text-independent speaker identification has been developed which utilizes vowel sounds as the basis for extracting speaker characteristics. The use of this approach typically requires that vowel samples first be classified according to vowel category, so that vowels of the same category can be compared in the speaker identification process. It has been demonstrated, however, that it is only necessary to detect vowel-like sounds in the speech material and that speaker identification performance actually improves when there is no vowel recognition. This eliminates an extra decision stage in the process, thereby reducing the computation load and the overall system error rate. An explanation for the improved identification performance is that the pooled vowel samples from each speaker are a representation of each person's vowel space, which is expected to be very speaker-dependent.

Another significant outcome of this research was the successful application of sequential analysis to the decision process. Sequential analysis relies on the accumulation of speaker classification results from several vowel samples before making a decision. If a sufficient number of test samples are classified as any one speaker, than a decision can be made, with a certain level of confidence, regarding the identity of the unknown speaker. The method allows acceptance and rejection thresholds to be established with specified error probabilities. This makes the sequential analysis procedure a dynamic process which accumulates and tests vowel samples until a confident decision can be made. Thus, the decision time and the amount of speech material needed is variable, depending on the speaker being tested. A speaker whose voice is distinct would be identified in a short period of time, whereas one whose voice is similar to others might require a larger number of samples. The sequential analysis is similar to the human perception process where we can quickly identify a unique voice, but listen longer when there is uncertainty.

The experiments which led to the conclusions of the research were based upon a very large data base of vowel samples from more than one hour of speech material excerpted from conversational speech recordings. Five vowel classes were represented in the data, these being among the most frequently occurring vowel sounds in spoken English. Recordings were made of twenty speakers (ten male and ten female) during conversational interviews. Two of the speakers were recorded two-weeks later so that effects of time could be studied as well. Three minutes of interview with each speaker was digitized and stored in computer files. The vowel samples were located and labelled in a semiautomatic manner with the assistance of an automatic boundary detection algorithm, which marked the location of potential boundaries between sounds. This provided a valuable assistance to the operator and resulted in a consistent criterion for delimiting vowel sounds. The most stable location of each vowel was also detected automatically by an algorithm which determines the location of least spectral change within a given speech segment. The autocorrelation method of linear prediction was used to analyze a 20msec window in the steady-state portion of each vowel. Twelve reflection coefficients (k-parameters) were generated and used as feature vectors throughout the study. A total of 4786 vowels were labelled and analyzed. These data were split into two independent sets for reference samples and test samples.

The first series of three experiments led to two significant results: 1) a majority-rule decision procedure could provide 95 percent correct identification, and 2) it is possible to have high accuracy speaker identification from vowel sounds without vowel recognition. It was found that if many vowel sounds from an unknown utterance are extracted and classified according to speaker, more samples will be classified as the true speaker than any other speaker, 95 percent of the time. Furthermore, when the vowels were not gouped into vowel categories, 45 percent of the samples were correctly classified (according to speaker), as opposed to only 39 percent when vowel categorization was performed.

Additional experimentation involved the use of sequential analysis in the decision process. Decision thresholds were established with the error specifications that P[FR]=.05 and P[FA]=.05. Using sequential analysis, 18 out of 20 speakers (90 percent) were correctly identified, 17 with the specified probability level, and one at a slightly lower confidence level. The amount of time required to reach a decision ranged between 5.5 seconds and 90 seconds. For one of the test speakers the decision narrowed down to two possible references, with the true speaker being a member of the pair. Only one test speaker in twenty was falsely identified.

Using the same 20-speaker reference patterns, tests with data from two speakers recorded two-weeks after the reference material showed that one of the speakers was correctly identified in 18 seconds and the second speaker was correctly identified, but with a lower confidence level than the first.

The results of this study have demonstrated new concepts which can be successfullly applied to advance the state of speaker identification. A completely automatic text-independent speaker identification is possible with the methods and procedures described in this report. Such a system could process an arbitrary speech utterance or recording in a sequential manner, and produce a decision as soon as possible, with specified levels of confidence. The following section lists and describes the recommendations for the additional effort required to develop a reliable automatic speaker identification system.

5.2 Recommendations

5.2.1 Automatic vowel detection

Since vowel recognition will not be required in the identification process, it is sufficient to have an automatic method of locating vowel-like segments in informal speech. Part of this procedure already exists in the form of the boundary detection algorithm used in this study. An additional level of intelligence needs to be added to the boundary program to make it decide if the speech segment between a consecutive pair of boundaries is a vowel-like sound. When a vowel-like sound is detected, the existing steady-state algorithm can be invoked to find the location for analysis.

An automatic vowel detection procedure would make it possible to collect large amounts of data since the much slower hand labelling process would not be necessary. Data from a large number of speakers, representing various recording conditions and over a range of time could be gathered in a reasonable period of time. This would permit exhaustive testing of various methods and procedures and provide statistically stable results for practical evaluation of true error rates.

5.2.2 Automatic sequential analysis

In this study, the sequential analysis procedure was successfully applied, but not as an automatic process. The sequential classifications were computed and then displayed, so that the decision stage was performed through a graphic interpretation of the sequential results. The automation of this process is merely a programming exercise.

Functions which should be included in the automatic sequential analysis are: 1) termination of a test when the acceptance threshold has been reached, 2) removal of references which have reached the rejection threshold, thereby reducing the reference population as the test progresses, and 3) if a decision has not been reached after a certain amount of time (or a certain number of samples) either force a decision or start over. Additional study may also be necessary on other decisions which can be handled with this procedure, such as imposter rejection.

5.2.3 Degradation due to time and channel

In the current study, degradation due to time was addressed on a relatively small scale. While the results were quite satisfactory, additional data over a longer period needs to be examined. The automatic vowel detection would make such an undertaking possible.

Degradation due to channel differences is an important aspect which should be addressed in future studies. Speech recordings over various communication channels should be included in a speaker identification data base. The speaker identification procedure should be tested using a statistical representation of the reflection coefficients e.g., the standard deviation computed over several vowel samples. This may help to normalize inter-channel variations and therefore reduce potential channel sensitivity.

5.2.4 Identification by verification

The basic assumption in most speaker identification studies is that there is a closed population of reference speakers, i.e., that it's a closed-choice test. This may not be realistic in most practical situations, however, so allowances should be made in the decision procedure for the possibility that the unknown speaker is not among the reference population. It is suggested, therefore, that speaker identification might be approached from a speaker verification point of view, with each reference being evaluated separately for similarity to the test speaker. Thus, if none of the references is sufficiently similar, then the test speaker can be rejected with identity unknown. This would result in a general purpose text-independent speaker identification/verification procedure.

SECTION 6

REFERENCES

Paul, J.E., Rabinowitz, A.S., Riganati, J.P., and Richardson, J.M. (1974). "Semi-Automatic Speaker Identification System (SASIS)-Analytical Studies Final Report", Rockwell International Report No. C74-1184/501, Prepared for the Aerospace Corp.

Aerospace Corp. (1977). "Speaker Identification, Program Final Report", Aerospace Report No. ATR-77(7617-07)-1.

Pfeifer, Larry L. (1977). "An Interactive Laboratory System for Research in Speech and Signal Processing", submitted to IEEE Transactions on Acoustics, Speech, and Signal Processing.

Wald, Abraham (1952). Sequential Analysis, John Wiley and Sons, Inc. New York.

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.

AMERICAN REVOLUTION BICENTENNIAL
1776-1976